

RESULT 9  
TCNA\_TRYCR STANDARD: PRT: 1162 AA.  
AC P23253;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 20-NOV-2001 (Rel. 40, Last annotation update)  
DE SIALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).  
GN TCNA.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RX SEQUENCE FROM N.A.  
RC STRAIN-SILVIO X-10/4;  
RA MEDLINE=91277609; PubMed=1711561;  
RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,  
RA Prioli R.P.;  
RT "The Trypanosoma cruzi neuraminidase contains sequences similar to  
RT bacterial neuraminidases, WTD repeats of the low density lipoprotein  
RT receptor, and type III modules of fibronectin."  
RL J. Exp. Med. 174:179-191(1991).  
RN (2)  
RP SUBCELLULAR LOCATION.  
RX MEDLINE=91376547; PubMed=1896773;  
RA Prioli R.P., Mejia J.S., Ajl T., Alkawa M., Pereira M.E.A.;  
RT "Trypanosoma cruzi: Localization of neuraminidase on the surface of  
RT trypanomastigotes."  
RL Trop. Med. Parasitol. 42:146-150(1991).  
CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN  
CC PARASITE INVASION OF CELLS.  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC  
CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYNEURAMINYL  
CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETATED  
CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,  
CC GLYCOLIPIDS OR COLOMINIC ACID.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
CC (POSSIBLE).  
CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPANOMASTIGOTES, MINIMUM  
CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.  
CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT  
CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TCNA  
CC PROTEIN.  
CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.  
CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC -----  
CC EMBL: M61732; AAA30255.1; -  
DR PIR: JH0557; JH0557.  
DR HSSP: P29768; IDIL.  
DR Interpro: IPR002860; BNR.  
DR Pfam: PF02012; BNR. 2.  
KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;  
KW Phosphorylation.  
FT DOMAIN 1 457  
FT REPEAT 23 34 CYS-RICH.  
FT REPEAT 163 174 BNR 1.  
FT REPEAT 209 220 BNR 2.  
FT DOMAIN 458 588 BNR 3.  
FT DOMAIN 589 1120 FIBRONECTIN TYPE-III.  
FT CARBOHYD 342 342 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.  
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 7.3%; Score 530.5; DB 1; Length 1162;  
Best Local Similarity 29.4%; Pred. No. 3.7e-14;  
Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;

250

QY 313 TPTTPEKE-PASTTPKEPTPTTTSAPTTPKEKPAPTTTKSAPTTPKEKPAPTTPPTTPKEKPAPTTP 371  
DB 599 TPSTPADSSAHSSTPVPVDS--SAHSTPSTPADSSAHSSTPVPVDSAGT--PSTP 651  
QY 372 KE-PAPTPKEKPAPTTTKSAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPA 418  
DB 652 ADSAHSAGTSTPVDSSAHSSTPVPVDSAGTSTPVDSSAHSAGTSTPVDSSAHSAGTSTPVDSSA 711  
QY 419 KEPTTP-TTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTP 477  
DB 712 VDSAGTSTPVDSSAHSSTPVPVDS--AHSSTPSTPADSSAHSSTPVPVDSAGT--PSTP 764  
QY 478 PTPKEKPAPTTTKSAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTP 533  
DB 765 HGTPSTPADSSAHSSTPST--PADSSAHSAGTSTPVPVDSAGT--PSTPVDSSAHSAGTSTP 817  
QY 534 TP-KKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTP 588  
DB 818 TPVDSSAHSSTPVPVDSAGTSTPVDSSAHSSTPVPVDSAGTSTPVPVDSAGTSTPVPVDSAGT 876  
QY 589 TTP-EKLAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTP 638  
DB 589 TTP-EKLAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTP 698  
APTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTP 698  
ADSSAHSSTPSTPADSS--AHSSTPSTPADSSAHSSTPSTPADSSAHSSTPSTPADSSAHSSTPSTPADSS 983  
3TAPTPKE-PAPTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTP 757  
PSTPADSSAHSSTPVPVDSAGTSTPVPVDSAGTSTPVPVDSAGTSTPVPVDSAGTSTPVPVDSAGTSTP 1033  
ISDKPAPTPK---ETAPTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPAPTTPKEKPA 813  
PSTPADSSAHSSTPVPVDSAGTSTPVPVDSAGTSTPVPVDSAGTSTPVPVDSAGTSTPVPVDSAGTSTP 1084  
IHSPDESSTP-ELSAETPAPALNSPREPVPPTTKPAATKPE 872  
DB 1085 DSSAH-STPSTPADV--AHGTP--STPADSSAHSSTP-----STPADSSAN 1125  
QY 873 MT 874  
DB 1126 GT 1127  
RESULT 10  
ID CPN\_DROME STANDARD: PRT: 865 AA.  
AC 002910;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE CALPHOTIN.  
GN CPN OR CAP.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
GN Ephydriidae; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CANTON-S.  
RX MEDLINE=93165729; PubMed=8094559;  
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;  
RT "alphoto: a Drosophila photoreceptor cell calcium-binding protein."  
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CANTON-S.

```

RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT calcium and contains a leucine zipper."
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC OF CA2 PER MOL. OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC COMPOUND EYES AND OCCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC DEVELOPMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L02111; AAA28405.1; -
DR EMBL; L05080; AAA28420.1; -
DR PIR; A47282; A47282.
DR FlyBase; FBgn0010218; Cpn.
KW Calcium-binding.
FT CONFILCT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFILCT 43 43 I -> T (IN REF. 2).
FT CONFILCT 64 64 I -> V (IN REF. 2).
FT CONFILCT 76 76 T -> A (IN REF. 2).
FT CONFILCT 100 100 P -> PP (IN REF. 2).
FT CONFILCT 126 127 VQ -> AP (IN REF. 2).
FT CONFILCT 154 154 I -> V (IN REF. 2).
FT CONFILCT 160 160 S -> T (IN REF. 2).
FT CONFILCT 534 534 A -> E (IN REF. 2).
FT CONFILCT 699 699 I -> T (IN REF. 2).
FT CONFILCT 703 703 V -> L (IN REF. 2).
FT CONFILCT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFC CRC64;

```

```

Db 357 AAESIPAPVATTPVPAT---LAVIDPVTASAVEPELPVIAAPSPVSAVAETPVDLAP 413
Qy 646 TPKEPAPTPKEPAPTPKEPAPTP- KGAPPTLKE-----PAPTPPKPADKE 694
Db 414 V----LPVAAEPPVAAVAAEETPETPAPASAPVIALMDIEVAVIAAPSDADAEAPSA 469
Qy 695 LAPTTKPTSTSDKPAPTPKGAPTP-----KEPAPTTKEP--ATTTKGAPTP 746
Db 470 AAPIVSTPP--TTSVPEPTTAPPAVPTPEPIDVSVLSAALETVPAPVEVTEVAAADV 527
Qy 747 TLKEPA-----PTPKKPAKEIAPTTT-----KGPSTSDKAPPT---PKETAPTT 792
Db 528 APPEAAADLIEPPEPPAPIDLEQTTSPVAAVEASTSS--PIPEISLPPMEAAVS 585
Qy 793 KEPAPTPKKPAPTTPPTPTTSEVSTPTTKEPTTIHSPD-----ESTPEL 842
Db 586 EVAVAPITAPPIP-----EPEPSLATPEPIPVAPVIOEAVAVAEVPTETSTSIPT 641
Qy 843 SAEPKPALENSPKPEGVPTTKTAPAKREMTTAKOKTTRDILRTTP--ETTAAPKMT 900
Db 642 TVE-FPEVAEKVLDPAI--TEAPVTOEPDVANINDGAPATEI--TPPAEIVTAAAEVS 697
Qy 901 KETATTEKTESKITATTTQVTSF-----TTQDTPPKITTLKTTLA--PKYTTTKT 953
Db 698 DIAIPVIDPVPQELAAVEIETDTKAEVIEGSTIPIEAPVPEVSKYAPVISEAPAA 757
Qy 954 ITTEIANKPEETIAKPKDRATNSKATTPPKOKPFPKAPKPTSTYKPTMPRYRKPTTPT 1013
Db 758 EVPTAGNDPNTSVGISSEVPTIAEKREVEVPTS--EIPQSSSPS--DSVPAKTIPL 813
Qy 1014 PRKMTSTPELNPSTSRIAEAM---LQTTTRPNQPNKSLVNVN-----PKS 1056
Db 814 LRDLOTIDVSLAATLAIATLAIIGELKDKOKARNOQVMORICETIKTLPPPS 864

RESULT 11
FPI_MYTGO
ID FPI_MYTGO STANDARD; PRT; 872 AA.
AC 025434;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCPPI).
GN FPI.
OS Mytilus coruscus (Sea mussel).
OC Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC NCBI_TaxID=42192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Foot.
RX MEDLINE=96394686; PubMed=8798340;
RA Inoue K., Takeuchi Y., Takeyama S., Yamahata E., Yamazaki F., Odo S.,
RA Harayama S.;
RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and
RT its evolutionary implications."
RT J. Mol. Evol. 43:348-356(1996).
CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC -1- PTM: PROLINES (THUS PRODUCING DOPA = 3,4-DIHYDROXYHENEYLALANINE).
CC ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYHENEYLALANINE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@slb-sib.ch).
CC -----
DR EMBL: D63777; BAA09850.1;
DR InterPro: IPR002964; Adhesive_plaq.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1216; ADHESIVE1.
DR PRINTS: PRO1217; PRICHTEXTENS.
KW SIGNAL; Repeat; Hydroxylation.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.
FT DOMAIN 21 41 NONREPEAT LINKER.
FT DOMAIN 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-
FT DOMAIN 184 192 P-[PS]-[ST]-Y-
FT DOMAIN 213 221 NONAPEPTIDE 1.
FT DOMAIN 213 221 NONAPEPTIDE 2.
SQ SEQUENCE 872 AA; 101677 MW; 98CC70D7C5FEFC4 CRC64;

Query Match 7.18; Score 518.5; DB 1; Length 872;
Best Local Similarity 29.24; Pred. No. 8.5e-14;
Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

QY 180 VYDAGSGLDNGDFKVTTPDST--TQH-----NKVSTPKTTAKPIPRESLPPNS 230
DB 29 VYGSAYSGASAGAYK-TLPSSHYPYSGAHVYKPMNKIPI-PIY--SKSYPAVYKPKGY 84
QY 231 DTSKETSLIVNKEETVETKETTTNKOTSDGKEKTSAKETQSIKTSADKLAPTSVL 290
DB 85 YPTNRVQPTGSKTNPPYKPIAKLSYKAITTPYAKAKSYPPSKY-----HKIT 139
QY 291 AKPTPKAETTKGPAITTPKEPT--PTTPKEPASTTPKEEPTTTIKSAPTTPKAPTTTK 349
DB 140 YPPPYK-----PKITYP--PYKOKPSYPSYKPKTYTPPYK-----PKITYPPYK 185
QY 350 SAPT-TPEKAPPT--TKKAPPTPKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPT--PK 404
DB 186 RKPSYTPYKPAITPPYKPKITYP---PYKPKPST---PYKPKTYPPYKPKI 236
QY 405 PAPTTPKEPAPT--PKEPTPTT--PKEPAPTTPKEPAPTTPKEPAPT--APKAPAPTTP 457
DB 237 SYSPYKPKASYSYSSKSKTKTYPPYKPKISYPPYKPKPSYPPYKPKTYPPYK 292
QY 458 KEAPPTPKKAPPTTKKAPPTTKKAPPTTKKAPPT--TKKAPPTTKKAPPTTK 512
DB 293 K-PKPSYPPYKPKITYPPT--YKPKPSYPPYKOKPSYPPYKSKSYPSYKSKTKYPP 350
QY 513 EPSPTTKKPA--PTTPKEPAPTTPKAP--APT-----EPAPTTPKAPAPTTPKKP 562
DB 351 TYKPKITYPPYKPKPSYPSYKPKTKYSPYKPKITYPPYKPKPSYPSYKPKTYPP 410
QY 563 APAPKEPAPTTPKEAP--TTPKKILPTTPKELAPTTTPKEPAPTTPPELAPTTPEE 617
DB 411 T-YKPKISYPPYKPKASYSYSSKSKTKTYPPYKPKISYPPYKPKPSYPPYKPKITYP 469
QY 618 PT-PTTPEEAPPTPKAAAPNTERK--EPAPTTPKEPAPTTP--KAPAPT--PKETAP 668
DB 470 PTYKPKSYPPYKPKITYPPYKPKPSYPPYKOKPSYPPYKSKSYPSYKSKTKYPP 529
QY 669 TT--PKGAPTTPKEPAPTTPKAPKELAPTTTKEPTSTSDKAPPT--PKGAPAPTTPK 725
DB 530 PTYKPKITYPPYK-----PKPSYPSYKPKTKTYPPYKPKITYPPYKPKASYPYK 583
QY 726 EPAPTTPKEPAPTTPKGTAPTTPKEPAPTTP--KKAPKELAPTTTGTSTSDKPA 781
DB 584 -----PKITYPPYK--PKPSYTPYKOKPSYPPYKSKSYPSYKSKTKYPP 628
QY 782 PTT--PKETAPTTPKEPAPTTPK--KAPAPTTPET-----PPTTSEVSTPTTKKEPT 830
DB 629 PPTPKPKITYPPYK--PKPSYPSYKPKITYPPYKPKKSKSYPOAKSKSGSPSYQPKKT 687
QY 831 IHKSPDSTELASPTPKALENSPKKPGVTTTPATKREEMTTAKDKT-----TENDL 886
DB 688 Y---PSPYKPKTKTYPPYK-----PKISYPPYK-----TKPSYASPKRKTSTYPTTKPK 735

```

```

QY 887 RTPEPTTAAPKMTKETATTEKTESKITATTTQVSTTTQDTPFKITTLKTTLAPK 946
DB 736 ISYSTYKAK-----SYPTPKPKPSYASSYKPKIRIPYKPKPSYASSYK-----PK 785
QY 947 VTTTKKTTTTTINKKPEETAKPKDRATNSKATTPKPO--KPTKAPKPP--TSTKPKK-T 1001
DB 786 I-----RPPYTKPKP-----SYASSYKPKIRIPYKPKPSYASSYKPKIT 827
QY 1002 MPYKPKKPTT--PPPKKTSMPPELNTPSRLAEMLQTTTPPNQ 1043
DB 828 YPTPKPKISYPTPKPKITYPPYKPKISYPPYKPKPSYQ 871

RESULT 12
NFH_RAT STANDARD; PRT: 831 AA.
ID NFH_RAT
AC P16884; 063368;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
GN NFH OR NFH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89065087; PubMed=1143606;
RA Breen K.C., Robinson P.A., Wion D., Anderston B.H.;
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT Identification of putative phosphorylation sites."
RL FEBS Lett. 241:213-218(1988).
RN [2]
RP SEQUENCE OF 37-831 FROM N.A.
RX MEDLINE=88309090; PubMed=2457365;
RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
RA Joles P.;
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT in situ detection."
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RN [3]
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX MEDLINE=87080760; PubMed=2878828;
RA Robinson P.A., Wion D., Anderston B.H.;
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT (NF-H)."
RL FEBS Lett. 209:203-205(1986).
RN [4]
RP SEQUENCE OF 318-831 FROM N.A.
RX MEDLINE=89184647; PubMed=2928342;
RA Lieberburg I., Spliner N., Snyder S., Anderson J., Goldgaber D.,
RA Smolowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
RT "Cloning of a cDNA encoding the rat high molecular weight
RT neurofilament peptide (NF-H): developmental and tissue expression in
RT the rat, and mapping of its human homologue to chromosomes 1 and
RT 22."
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -I- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND

```





DR	HSSP; P04002; 1WPA.
DR	InterPro; IPR000519; P_trefoll.
DR	Pfam; PF000086; trefoll; 6.
DR	SMART; SM00018; P_6.
DR	PROSITE; PS00025; P_TREFOLL; 6.
KM	Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
FT	NON_TER
FT	DOMAIN
FT	81
FT	144
FT	REPEAT
FT	81
FT	88
FT	89
FT	96
FT	97
FT	104
FT	105
FT	112
FT	113
FT	120
FT	121
FT	128
FT	129
FT	136
FT	137
FT	144
FT	DOMAIN
FT	161
FT	202
FT	218
FT	301
FT	REPEAT
FT	218
FT	224
FT	225
FT	239
FT	240
FT	249
FT	250
FT	259
FT	260
FT	275
FT	REPEAT
FT	276
FT	287
FT	REPEAT
FT	288
FT	294
FT	REPEAT
FT	295
FT	301
FT	DOMAIN
FT	306
FT	347
FT	353
FT	394
FT	DOMAIN
FT	402
FT	522
FT	REPEAT
FT	402
FT	411
FT	412
FT	419
FT	REPEAT
FT	420
FT	431
FT	REPEAT
FT	432
FT	443
FT	REPEAT
FT	444
FT	453
FT	REPEAT
FT	454
FT	460
FT	REPEAT
FT	461
FT	472
FT	REPEAT
FT	473
FT	479
FT	REPEAT
FT	480
FT	491
FT	REPEAT
FT	492
FT	498
FT	REPEAT
FT	499
FT	515
FT	REPEAT
FT	516
FT	522
FT	DOMAIN
FT	525
FT	566
FT	DOMAIN
FT	572
FT	613
FT	DOMAIN
FT	620
FT	661
FT	DISULFID
FT	162
FT	188
FT	DISULFID
FT	172
FT	187
FT	DISULFID
FT	182
FT	199
FT	DISULFID
FT	307
FT	333
FT	DISULFID
FT	317
FT	332
FT	DISULFID
FT	327
FT	344
FT	DISULFID
FT	354
FT	380
FT	DISULFID
FT	364
FT	379
FT	DISULFID
FT	374
FT	391
FT	DISULFID
FT	526
FT	552
FT	DISULFID
FT	536
FT	551
FT	DISULFID
FT	546
FT	563
FT	DISULFID
FT	573
FT	599
FT	DISULFID
FT	583
FT	598
FT	DISULFID
FT	593
FT	610
FT	DISULFID
FT	621
FT	647
FT	DISULFID
FT	631
FT	646
FT	DISULFID
FT	641
FT	658
FT	VARIANT
FT	276
FT	276
FT	VARIANT
FT	354
FT	354
FT	VARIANT
FT	415
FT	415
SO	SEQUENCE
662 AA;	67774 MM;
F085277F1ED2FD40 CRC64;	T -> A.
	C -> R.
	A/A/THR-RICH.
	8 X 8 AA APPROXIMATE TANDEM REPEATS,
	ALA/THR-RICH.
	1-1.
	1-2.
	1-3.
	1-4.
	1-5.
	1-6.
	1-7.
	1-8.
	P-TYPE 1.
	8 X APPROXIMATE TANDEM REPEATS, THR-RICH
	2-1.
	2-2.
	2-3.
	2-4.
	2-5.
	2-6.
	2-7.
	2-8.
	P-TYPE 2.
	P-TYPE 3.
	12 X APPROXIMATE TANDEM REPEATS,
	THR-RICH.
	3-1.
	3-2.
	3-3.
	3-4.
	3-5.
	3-6.
	3-7.
	3-8.
	3-9.
	3-10.
	3-11.
	3-12.
	P-TYPE 4.
	P-TYPE 5.
	P-TYPE 6.
	BY SIMILARITY.
	BY SIMILARITY.
	BY SIMILARITY

Query Match	7.0%;	Score 509.5;	DB 1;	Length 662;
Best Local Similarity	28.1%;	Pred. No. 1.5e-13;		
Matches 223; Conservative	48;	Mismatches 276;	Indels 247;	Gaps 30;

QY	253	TTTKKQSTKQSEKETTSAKESQISAAADLAPTSKVLAKPPKKAETKTTKCPALUPTTKEP	312
Db	3	TTTAAANAANTKDTTAAEGSAAAEKTA-----AGEVSNPPT--AAANAATEDATTT-----	51
QY	313	TTPTTPEKPASTTKEEPTPTTIKSAPTTPEKPAPTTTSKAPTTTPEKPAPTTTSKAPTTTPEK	372
Db	52	-----AAATAAAEETTAAGCAPTTTTATTAAGAPTTTAAATATTTAAAGAPTTTAT	103
QY	373	EPAPTTTKEPAPTPTTYS-APTTPEKAPPTTTPKRAPTTPEKAPPTTKEEPTPTTPEKPA-	430
Db	104	GKAPATTAAPVPTTAAASKAPTTAAATHTSTAAAPPAATTAASAKSEKSTSSSEEHCH	163
QY	431	-----PTTKEPAPT-----TPEKAPATPAKRAPPTT-----BEAPTPEKAPPTTKE	475
Db	164	VKPSRKMCGSKQITKQ-----CKKNCCPDKGGHGHCHFRKKKGSHSEHTTTTTK-	218
QY	476	PSPTPEKAPPTTTSKAPTTTKEPAPTTSKAPTTTKEPSPTTPEKAPPTTPEKAPPTT	535
Db	219	-----APTTIQATTTTT--PTTT-----TTTTKAPT-----	245
QY	536	KKRAPTTPEKAPPTTPEKAPPTTTPKRAPTAPEKAPPTTPEKTAAPTTPPKLLPTTPEKLA	595
Db	246	-----TTTTTKATPTT-----TTTTKATTTTT--TPTT-----	270
QY	596	PTTPEKAPPTTPELAPTTPEEPTPTTPEBAPTTPPKAAAPTPEKAPPTTPEKAPPTT	655
Db	271	-----TTTTTKATPTTPTTTTTT-----	289
QY	656	KEPAPTTPKEAPTTPEKGTAPTTTLEKAPPTTTPKPAKELAPTTTKEPTSTTSKAPATT	715
Db	290	-----TTTTTKATTTTTTTTSECKMEPS-----REDCGVSGLTESQCR	328
QY	716	PKG-----TAPT-----PEKAPPTTPEKAPPTTPEKGTAPTTTLEKAPATTPKRAPKELA	765
Db	329	TKGCCFDSISIPQTKWCFFYTLISQVADCKVPSQKRVDCGFGGIT-----ADCCRKKNCCFDS	384
QY	766	PTTTKGTSTTSKAPAPTTTPEKAPTTPEKAPATTPKKAAPTTPENTPPTTSEVSTPTT	825
Db	385	ISGKWCFFYSTQVAA--TCTTTPTTTTTPTTTTTTKATTTTTT-----TTTTTPTTT	436
QY	826	KEPTTIHKSPEDESTPELSAPTEPKALENSPEKEGVPPTTKPATKPEMTTAKDKTERD	885
Db	437	TTTTT-----TTKA-----TTTTPTTTPTTTTTT--KAT--	463
QY	886	LRTPEPTTAPKKTATTTTEKTESKITATTQVSTTQDPTPKITTLKTTTLAP	945
Db	464	TTTTTTTTT-----TTTTTKAT-----TTTTPTTTPTTTTTTKATTTTTPTTTTT	510
QY	946	KVTTTKTTTITTEIMNKPEETAPKPKRATNSKATTPPKPKPAKPKPTSTKPK-----T	1001
Db	511	TTTTTKATTTTTTSECKME-----PSKRADCGYTGITRESQCRKGCDFDSISIPQTKWCYS	566
QY	1002	MPRVAKPKPTTTPR	1015
Db	567	LPOVADCKVAPSSR	580
RESULT 14			
RPBL_HUMAN	STANDARD;	PRT;	1970 AA.
AC	P24928;		
DT	01-MAR-1992 (Rel. 21, Created)		
DT	01-MAR-1992 (Rel. 21, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPBL).		
GN	POLR2A.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	[1]		
SEQUENCE FROM N.A.			

```

RX MEDLINE=92178992; PubMed=1542581;
RA Watzlenth M., Acker J., Vicaire S., Vigneron M., Keding C.;
RT "Complete sequence of the human RNA polymerase II largest subunit.";
RL Nucleic Acids Res. 20:910-910(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=95347616; PubMed=7622068;
RA Mita K., Tsuji H., Morimyo M., Takahashi E., Neno M.,
RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
RT "The human gene encoding the largest subunit of RNA polymerase II.";
RL Gene 159:285-286(1995).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63564; CA45125.1; -
DR EMBL: X74874; CA52862.1; -
DR EMBL: X74873; CA52862.1; JOINED.
DR EMBL: X74872; CA52862.1; JOINED.
DR EMBL: X74871; CA52862.1; JOINED.
DR EMBL: X74870; CA52862.1; JOINED.
DR PIR: S21054; S21054.
DR MTM: 180660.
DR InterPro: IPR000684; RNA_polII_repeat.
DR InterPro: IPR000722; RNA_pol_A.
DR InterPro: IPR002879; RNA_pol_A2.
DR Pfam: PF00623; RNA_pol_A; 1.
DR Pfam: PF01854; RNA_pol_A2; 1.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; 43.
KW Transferrase; DNA-directed RNA polymerase; Transcription; zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.
FT ZN_FING 71 87 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT DOMAIN 1590 1958 C2H2-TYPE (POTENTIAL).
FT CONFLICT 1067 1067 W -> L (IN REF. 2).
FT CONFLICT 1449 1449 D -> Y (IN REF. 2).
FT SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;
SQ

```

Query Match 6.98; Score 503.5; DB 1; Length 1970;  
 Best Local Similarity 33.38; Pred. No. 6.4e-13;  
 Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

```

QY 312 PPTTPKEPASTPKE-PPTTIKSAPTTKEPAPTTSKAPTTPKE-----PAPTTPKE 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1507 PSPMGISPMATPMNQCATRAYGAMSPSGMTPGAGSPSASASDASGSPSYSPAMS 1566
QY 366 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKAPPTTPKEPAPTTPKEPPT 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1567 PTPGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1624
QY 424 TTPKEPAPTTPKEPAPTTPKAPPTTPKEPAPTTPKAPPTTPKAPPTTPKAPPTTPKE 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1625 TSPTSP-SPTSPNYSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1671

```

```

QY 484 PAPTTPKSAPTTPKAPPTTPKSAPTTPKESPTTPKEPAPTTPKEPAPTTPKAPPTTP 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1672 -SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1725
QY 544 KEPAPTTPKEPAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPT 603
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-T 1777
QY 604 PTPPELAPPTPEEPPTTPPEEPAPTTPKAAPNPKAPPTTPKEPAPTTPKEPAPTTPKEP 661
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1778 PTPSP-NISPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1828
QY 662 TPKEPAPTTPKGAAPTTPKAPPTTPK-KPAKELAPPTTPKEPSTSDKAPPTTPKGA 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1829 SP-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1875
QY 721 PTPKEPAPTTPKAPPTTPKGAAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPT 779
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1876 PTPSPY-SPTSPY-SPTSPY-SPTSPY-SPTSPY-SPTSPY-SPTSPY-SPTSPY-SPTSP 1922
QY 780 PAPTTPKAPTTPKEPAPTTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1923 YSPTSP-TYSPSPKGSYSPSPSY-SPTSP-----YSPTSPAL-----SPDD 1965
QY 838 STPE 841
DB 1966 SDEE 1969

```

```

RESULT 15
NFH_HUMAN STANDARD; PRT; 1020 AA.
ID P12036;
AC 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN NFH OR NFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328981; PubMed=3138108;
RA Lees J.F., Smeidman P.S., Skuntz S.F., Carden M.J., Lazzerini R.A.;
RT "The structure and organization of the human heavy neurofilament
RT subunit (NF-H) and the gene encoding it.";
RL EMBO J. 7:1947-1955(1988).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC OBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:49 ; Search time 61.21 Seconds

(Without alignments)  
502.197 Million cell updates/sec

Title: AAF  
Perfect score: 7294  
Sequence: 1 MAWKPLPYLLLLSVEVIQ.....ARATTRSGQLSKWYNCP 1366

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCMTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pcp: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522	7.2	805	4	US-09-103-429A-4
2	513	7.0	1837	3	US-08-928-361B-5
3	506.5	6.9	744	6	5202236-25
4	499.5	6.8	1721	3	US-08-700-651-5
5	498.5	6.8	766	4	US-09-103-429A-3
6	498.5	6.8	1721	3	US-08-928-361B-6
7	488.5	6.7	826	1	US-07-638-431-2
8	488.5	6.7	826	5	PCT-US97-00018-2
9	476.5	6.5	652	6	5202236-13
10	462	6.2	960	4	US-09-219-849-5
11	425.5	5.8	1185	4	US-09-041-886-23
12	424.5	5.8	1867	2	US-08-479-537A-5
13	424.5	5.8	1867	4	US-09-083-116-5
14	424.5	5.8	2035	2	US-08-479-537A-2
15	424.5	5.8	2035	4	US-09-083-116-2
16	419.5	5.8	2476	2	US-08-276-967-2
17	417	5.7	829	1	US-08-642-255-132
18	417	5.7	829	1	US-08-397-633A-53
19	417	5.7	837	1	US-08-175-155-68
20	417	5.7	837	1	US-08-477-509B-103
21	417	5.7	837	1	US-08-642-255-101
22	417	5.7	837	2	US-08-707-237A-75
23	417	5.7	837	3	US-08-482-085B-103
24	417	5.7	837	1	US-08-397-633A-50
25	413.5	5.7	907	5	PCT-US95-04611A-19
26	413.5	5.7	907	5	PCT-US95-04611A-19
27	404.5	5.5	408	1	US-07-609-716-65

28	404.5	5.5	408	4	US-08-475-411A-65	Sequence 65, App1
29	404.5	5.5	408	4	US-08-478-029A-65	Sequence 65, App1
30	398.5	5.5	682	1	US-08-642-255-126	Sequence 126, App1
31	398.5	5.5	682	1	US-08-397-633A-36	Sequence 36, App1
32	387	5.3	1537	1	US-08-325-267A-2	Sequence 2, App1
33	383.5	5.3	1231	3	US-08-904-263A-4	Sequence 4, App1
34	381	5.2	1848	5	PCT-US95-10661A-6	Sequence 6, App1
35	381	5.2	1848	5	PCT-US95-10661A-6	Sequence 6, App1
36	376	5.2	960	4	US-09-219-849-6	Sequence 6, App1
37	373	5.1	1786	4	US-08-973-462-8	Sequence 8, App1
38	371	5.1	761	2	US-08-707-237A-84	Sequence 84, App1
39	371	5.1	762	1	US-08-642-255-114	Sequence 114, App1
40	371	5.1	762	1	US-08-397-633A-26	Sequence 26, App1
41	370.5	5.1	1064	1	US-08-642-255-62	Sequence 62, App1
42	370.5	5.1	1187	1	US-08-320-559-28	Sequence 28, App1
43	370.5	5.1	1187	5	US-08-545-860D-28	Sequence 28, App1
44	370.5	5.1	1187	5	PCT-US94-04496-28	Sequence 28, App1
45	370.5	5.1	1210	1	US-08-320-559-26	Sequence 26, App1

## ALIGNMENTS

```
RESULT 1
US-09-103-429A-4
: Sequence 4, Application US/09103429A
: Patent No. 6187558
: GENERAL INFORMATION:
: APPLICANT: Granados, Robert R
: APPLICANT: Wang, Ping
: TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucln
: TITLE OF INVENTION: CDNA and Related Products and Methods
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Brown, Pimisi & Michaels, P. C.
: STREET: 118 No. 6187558th Tloga
: CITY: Ithaca
: STATE: NY
: COUNTRY: USA
: ZIP: 14850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/103.429A
: FILING DATE: 24-JUN-1998
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Michaels, Christopher A
: REGISTRATION NUMBER: 34,390
: REFERENCE/DOCKET NUMBER: BTI-39
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (607) 256-3628
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 805 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Trichoplusia ni
: TISSUE TYPE: peritrophic membrane
: US-09-103-429A-4
```

Query Match 7.2%; Score 522; DB 4; Length 805;  
Best Local Similarity 27.8%; Pred. No. 5.8e-26;

Matches	226; Conservative	33; Mismatches	211; Indels	342; Gaps	40;
OY	286	TSKVLAKPTPKAEATTGKGPALTTKEPRLPTTPKEPASTTPKKEPTPTTKSAP-----	TTP	340	
Db	9	TALGLVAARPEVSDAEKNPDLHEPHDPX--PAEQXXLLPXYDCTKFEYCEYGLAKFIAP		66	
OY	341	KEPAPTT-----TKSAPTTKEPAATTTTKEPAATTPKKEPAATTPKKEPAATTPK		388	
Db	67	RDCAGTLEPKFSAGQCVNAALAGCTLPGPRAETT---QAPATTO--APTTTT--APTTTT		119	
OY	389	SAPPTPKKEPAATTPKPKPAATTPKKEPAATTPKKEPTTPKKEPAATTPKKEPAATPA		448	
Db	120	QAPTTTTQATTTT-----QAPTTQAPTTT---QATTTQAPTTTTT-----		156	
OY	449	EKKRAPPTPKKEPAATTPKKEPAATTTKESPTTPKKEPAATTTKSAPPTTKKEPAATTKSAP		508	
Db	157	--QAPTTTTQ--APTTTTQ--APTTTTQ-----APTTTTQ--APTTTTQAA		198	
OY	509	TTPKPSPTTTPKKEPAATTP--KEPAATTPKPKAPATTPKKEPAATTPK--		552	
Db	199	TTPATTTTAAAT--TPAATTPAATTPAATTPGVAAPS---AAVWPRICGLLLNGCQADBD		253	
OY	553	-----	EPATTTK	560	
Db	254	IHLIPHDKYYCNLFYQCSNGYTFEORCEGLYFNPNYORCDSPANVECDGEISAPRPVE		313	
OY	561	K-----PA-----		563	
Db	314	GNEDEDIDIGLLDNGCFANFEIDMLPBGNCCKYQCVHGNLVBRGACGHSEFEO		373	
OY	564	-----	PTAP--KEPAT	578	
Db	374	QCDHIELVGCTLPGGESEVEDDADACTGWCPTPEIEMERLPNGCQADFSIDHLPHSES		433	
OY	579	-----APTTKXKLPLPT-----PKLAPPTPKKAP		604	
Db	434	DCGQYLQCVHGOTIARPCGNLHFSATQSCSPVTAGCQVECDSDNOCTSTAAPTAAP		493	
OY	605	TTPEELAPTTPEEPRTTPEEBEPATTPKAAAENTPEKEPATTPKKEPATTPKKEPATTPK		664	
Db	494	TAAPTAAPTAAPTAAPTASIVPRPA--TPPATAAAPVPPTTAIPT---PAPTAAPTAAPTTAA		548	
OY	665	ETAPPT---PKGTAPTTLKEPATTPPKKAPKELATTTKEPTSTSDKPAATTPKGTA		721	
Db	549	PESPTTIVVPPTAAPPA---APPT---AVDEIPTIVSAPTAAPT---AAPTAAPTAAP		598	
OY	722	TTPKPEAPPTPKKEPATTPKGTAPTTLKEPAATTPKPKAPKELAPTTKGPTST--TSDAP		780	
Db	599	TTAAVEIPTT-----VTSPTTAAPT--AAPAANT-----IVTVPTAAPTAAPTAAP		641	
OY	781	APTTKETAATTPKKEPATTPPKKPAATTPETRPPTTSEVSTPTTKKEPTTIKHSDESP		840	
Db	642	AP-----NTVTVAAPTAAPTT--AAPAANTTVTVPTAAPTAAAP-----PTVAH-----AP		685	
OY	841	ELSAEPTPKALENSPKREBGPVTTKTPAATKPE	872		
Db	686	NTTAAP-----VTTTSAPATT--PE	703		

CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94306-1840  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/928,361B  
 FILING DATE: 12-SEP-1997  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/026,062  
 FILING DATE: 13-SEP-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: VERNY, Hana  
 REGISTRATION NUMBER: 30,518  
 REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-324-1677  
 TELEFAX: 650-324-1678  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1837 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-928-361B-5

[illegible]

```

QY 538 ----- 537
Db 570 MIPNDTHVRFRKVDYNTISVRCGAGKLEPPDRSLDFTIPVAGHNSCIIVGS 629
QY 538 -----PARTPKE-----PARTPKEPARTTKK 561
Db 630 GGGRIHWSPYGSKDVSLISAPIQCELENEYCDCTAKYGAISHGYQTSADFTVTTAK 689
QY 562 PAPAPKPEAPPTTKEAPPTPKLPTTPEKLAPTTPEKAPPTPEAPPTPEEPTT 621
Db 690 PTTT---TGAPGQPTTTTGSFSPKPTTTTKATTT---TILNDITTTTQKPTT 741
QY 622 TPEEPAPTPKAAAPNTPEKAPPTTPEKAPPTTPEKAPPTTPEKAPPTTKE 681
Db 742 T-----TVGKPPPIATTTTTLKPIVTTTTKATTTTTPPT---TTTTKR 787
QY 682 PAPTPPKPAK---ELAPTTKEPTS-----TSDKAPATTGK-TAP 721
Db 788 DEMTTTTPLPDIGDIDITPIPIEKMLDKYRMIVDYNGLLSDNDEPIPGSAGQIAD 847
QY 722 TTPKEPA-----PTTPEKAPPTPK----- 741
Db 848 TSNLFPVQTHKSTGLPLDPMWGLPDPKSGNLVPIYTNQTMSCISVYLAKNLTVDIDE 907
QY 742 --GTAPTTLK---EPAPTPPKPAKPELAPTT-----KGPSTTSKDP 780
Db 908 TYGPIIDLGLGYPLDPVSLIPFNPELGELFDPISEIMNGTIAGIVSGISASELSOKS 967
QY 781 APTTPKE-----TAPTPKEAPPTTPKPAPTTPTP-----PPTTSVSTPTT 824
Db 968 APIDPATNMVGEFGLNPATGYMIGSLGPSQTFSEPIEDGGIIP---EVAANA 1024
QY 825 TKEPTTIHKSPPDESTEP-----LSAPTPKALENSPKKEGVP 861
Db 1025 DKFSLSTPSPVESIPEKDKKIDISISLMDIESGRILGOVSKRPFGISIGDLN---P 1080
QY 862 TTKTPAATKPEMTTAAKDKTERDLRTPTTTAAPKMTETA---TTEKTTESKITA 917
Db 1081 IMKPTPT-----DSVTKGPI---DPTTGLP-FNPTGHLINPTNNTMDSFAG 1126
QY 918 TTTQVSTTTQDTTPFKITTLTKTTTLAPK---VTTKKTTITTELMNKPPEEAKKDAT- 974
Db 1127 AYKAVSNGIKTDNVYGLPDEITGL-PKDPVSDIPNSTGELVD---PSTGKRIINNYTA 1183
QY 975 -----NSKATTPKPKQPTKAPKKEPTSKKRTMPRVKRPKPTTPPKMTS 1019
Db 1184 GIYSKRGRLPIEDENGNLDPSTKLPIDGNQNLVNPETNSTVSGTSKRPKPIPVN 1243
QY 1020 ---TME-----LNPISRIAEAMLTTRPNQTPNSKLVENPKSEDAGG 1061
Db 1244 GGGVVPDEAKDAQDKGKGLIIVPTNSINKDPVTNTQVSNITGNI---INP-ETGKV 1297
QY 1062 AEGEPTMLRPHVMEVTPD-----MDVLP RVNNGIINP----- 1099
Db 1298 IPGSLPESLWVPSFNTQOIDEITGKRVDTVYGLPIYPSGELIIDPRTKCLPIPSVAGDE 1357
QY 1100 MLDSEINICNG-----PVDGLTTLRNGTLV 1125
Db 1358 ILTEVLNITTDVYGLPIDLETGLPRDPVSGLPQLPGLTIV 1398

RESULT 3
5202236-25
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456

```

```

; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 25
; LENGTH: 744
; 5202236-25

Query Match 6.9%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 5.2e-25;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 212 PKTTAPINRPSLPPNSDTSKETSITVKNKETTVEKETTNNKQSTGDEKETTSAKE 271
Db 24 PKMTYPPYKPKPSYP-----YVSKSPY-----KPKIT----- 53
QY 272 TOSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPE---PASTPKE 327
Db 54 -----YPTTK-AKPS-----YPTTKPKTYPTTKLTPPTTKPKP 92
QY 328 PTPPTIKSAPT-TPKEBAPTTTGSAPTTPKBAPTTTKEBAPTTTKEBAPTTTKEBAP 384
Db 93 SYPTTKSKPTTKKITYPTTKAKPSYPTTKKPKTYPTTKKLTPTTKPKASYP 151
QY 385 TTTTSAPTTKEBAPTTPKKAPPT-TPKEBAPTT---PKBPTTTPKEBAPTTKEP--- 436
Db 152 PTVKPKPSYP--PSYTKTKTYPTTKKLTTPPTTKPKPSYPSYKPKTYPTTKKLT 209
QY 437 -APTTPKEPA--PTAPKPA--PTTPEKA--PTTPEKA--PTTKEPS--PTTPEKA 485
Db 210 YPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 269
QY 486 -PTTKSAPTTTKEBAPTTTKSAPTTKEBSPTTTKEBA--PTTPEKA--PTTPEKP 538
Db 270 YPSYTKAKPSY--YPSYTKAKPTYP---PTKAKPSYPTTKAKPTYKAKPTYKAKP 322
QY 539 A--PTTPEKA--PTTPEKA--PTTTKKPA--PTAKKEBAPTTTKEBAPTTTKEBAP 590
Db 323 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 382
QY 591 PEKLAPTTPE--KPAPTTPEELAPTTPEEPT--PTTPEKA--PTTPEKA--AAPNT---P 639
Db 383 TYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 438
QY 640 KEPAPTTKEBAPTT---KPEBAPTTKETAPTT--PKGAPTTLKEBAPTTKPAK 693
Db 439 SYPTTKPKISYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 498
QY 694 ELAPTTTKEPTTSDKAPATTGTAPTTPEKAPATTPEKAPATTGTAPTTLKPEBA- 752
Db 499 SYPTTKAKPTTSTYKAKPTTK--AKPSY--PTTKAKPS-----YPTTKAKPSY 546
QY 753 -PTTTPKPAKELAPTTTGP--STTSDKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 807
Db 547 PPTTKAKPTTK-AKPTTKAKPTYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 601
QY 808 ---TPETPPTTSVSTPTTKE---PTTIHNSPD-----ESTPELSAETP-PKALENS 854
Db 602 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAK 661
QY 855 PKBGPPTTKPA-----TKPEMTTAAKDKTERDLRTPTTETTTAAP-----KMT 900
Db 662 PSYP--PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 719
QY 901 KETATTEKTTESKITATTT 920
Db 720 YKAKPTYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 739

RESULT 4
US-08-700-651-5

```

```

; Sequence 5, Application US/08700651B
; Patent No. 6013882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUY, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721.
; TYPE: PRN
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-5

```

```

Query Match      6.8%; Score 499.5; DB 3; Length 1721;
Best Local Similarity 21.8%; Pred. No. 4e-24;
Matches 300; Conservative 94; Mismatches 510; Indels 469; Gaps 44;

```

```

QY 93 STTKSRPKPKKKKKKVVIESEITEHVSSENOSSSSSSSSSTIWKIKS-----146
DB 38 SITGSRKDPKPKAT-----ISGRSCGKQGYSIDSGFVDSITGLPT 83
QY 147 --SKNSAANRLQKLKLVKDKKRNKTKKPPKPPVDEASGLDNGDFKTTTDTSTQ 204
DB 84 DPYSCNCFN-----PVTGLVSRSGTKTIPN-----TYAGVRSN-ETKTEPSANT--129
QY 205 HNKSSTPKITTAKP-----INPRSLPNSDTSKE-----TSLTVN 241
DB 130 YAGYRSNEMETTEPSANTNINLVDPKINAPCENSEFQOQIDMSKVYIPTKCGV 189
QY 242 KETTVETKETTNNKOSTDGEKETSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETT 301
DB 190 KHTTTTTTTTTTTTTTTTTT-----TTT 211
QY 302 KGRPLTPKKEPTPKKEPASTPKKEPTTISKAPTTPKKEPAPTTPKKEPAPT 361
DB 212 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 271
QY 362 TTKEPAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT 421
DB 272 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 331
QY 422 TPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 481
DB 332 TTTTTPKKTPTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTT 377
QY 482 KEPAPTTPKSAPTTPKKEPAPTTPKSAPTTPKE-----513
DB 378 TKPTTTTTTTTTTTTTTKKTTTTTATTTTTTSETSEVIKPDCEMCKNGDEAKATYVG 437
QY 514 -----PSPPT-----TKE 521
DB 438 VIGKGRIRIENGMAFTMINDTTHVRFKRVKDVGNTISVRCRGAGKLEFPDRSLDPTIP 497
QY 522 P-----APTTPKEPAPTTPKKEPAPTTPKE-----545
DB 498 PVAGHNSCIIVGVSGDGKIHVSPGSKDVSLIS---APIQPSLENEVYCDTCTAYG 553
QY 546 ---PAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 602
DB 554 AIHSQYQISAPFVTTTTTKKPTTTT--TGAPOPTTTTGGSSKPTTTTTTATTTTT 607
QY 603 APTTPEELAPTTPEEPTPTPEEPAPTTPKAAAPNTPREPAPTTPKKEPAPTTP 662

```

```

DB 608 --TLNPIIITTTOKPPTTT-----TT-----KVAGKRPPIATTTTLKPIYTTTTTKAFTT 656
QY 663 PKETAPTTPKGTAPTTLKEPAPTTPKPKAPK-----ELAPTTPKKEPT-----705
DB 657 TTTTVPPTT-----TTTTKREMTTTTTPLDIDGIEITPIPIEKMLDKRMIYDVSGL 711
QY 706 --TTSDBAPPTTPKG--TAPTTPKEPAPTTPKEPAPTTPKGTAPTLKE---PAPTTPKPK 759
DB 712 LIDSNDPEIPGSOAGQIADISNLFVOTHKSTGCLPIDPMVGLPDRKSGNLVHYTQTM 771
QY 760 APKELAPTTKGPSTTSDBKAPPTTPKEPAPTTPKEPAPTTPKPKAPPTTPETPTTSEV 819
DB 772 SGLSVSLAAKNLIVDIDETYG--LPIDITLGYLDPLVSLIPFN--PETGELFDPISDEI 827
QY 820 STPT-----TKEPTTIHKS-----PDST- 839
DB 828 MNGTIAGIVSGISASESLISOKSALIDPATNMVVGEGGLINPATGVIMPGELPSSDQTQ 887
QY 840 -----PULSAPPTPKALENSPKPEGVPTTKTPAATKPEMTTAKDKTTER---884
DB 888 FSPLEIDGGIIPPEYAAANADKFLSLIP--PSVP-----ESLPE-----KDKIDISIE 934
QY 885 -----DLRTPETTTAAPKMKREIATTEKTESKIYATT--TQVSTTTQD 929
DB 935 LMVDESGLIGVSKRPIPGSISAGDLNPIKPTQIDSVYGKPIDPTTGLPFNPBGHL 994
QY 930 TTPPKITTLKTTTLAPKYTTTKITTTTEIMNKP--EELAKPKO-----RATNSKAT 980
DB 995 INPNNMTMSSSFAGAAKVAVNSGICITDNVGLGVGELTGLPKDGSIDPNSITGELVYD 1054
QY 981 PKQPKTPAKPKPTSTKPKTMP-----RVKPKTTPPKRMTS 1019
DB 1055 PSTGKPIINSTAGIVSGKPGLPRIEDENGNLFDSTNLPIDGNQNLVNPENSVSGSTS 1114
QY 1020 --TME-----LNPTSLAEMLQTTTPPNOTPNSKL 1049
DB 1115 GTPKPKGIPVNGGCVVPEDEAKDQADKGDGLIVPPTNSINKDPVNTQVSNSTGNT--1172
QY 1050 VEVPKSESDAGCAGEMPHMLLRPHVMEVYTPD-----MDYLPVPNOGIIINP--1099
DB 1173 --INP--ETGKVIPIQSLPGSINTPSFNTPQOTDELTKGVPVTYVGLPDPSTGELIDPAT 1228
QY 1100 -----MLSDETINICNGK-----PVDGLTTLNGLTV 1125
DB 1229 KLPVGSVAGDELLTEVLNITTDVETGLPIDLETGLPRDPVSGLPQJLNGTIV 1281

```

```

RESULT 5
US-09-103-429A-3
; Sequence 3, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Nucin
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pimnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tiooga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998

```



361

```

RESULT 7
US-07-638-431-2
; Sequence 2, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumsmith, Rishin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: A. David Spevack
; STREET: NMRC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638.431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-638-431-2

Query Match 6.7%; Score 488.5; DB 1: Length 826;
Best Local Similarity 26.4%; Pred. No. 8.5e-24;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34

329 PPTTKTSAPPTTKEPEA-PPTTKSAPPTTKEPEA-PPTTKEP---APTTPKEP-APTTPKE 381
|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
273 TCKVADCDQIILPIPIPKIKIEKPSNPEEPVNPNDPNPNPNPNPNPNPNPNPNPN 332
382 PAPTTKTSAPPTTKEP-APTTPKP-APTTPKEP-APTTPKEP-PPTTKEPEAPTTPKEA 437
| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
333 PNNPNPNPNPNPNPNNNNNNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNNNHKK 392
438 PTTTPKEAPATJAK--KPAPTTTPKEPAAPTTPKEAPTTTKEPSPTTKEPEAPTTPKSAPTT 495
| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
393 RNNPKRRNNPKKPKPNKPNKPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 443
496 TKEPAPTTTSAPPTTKEPSPTTKEPEAPATTPKEA-PPTTPKP-APTTPKEAPPTTKE 553
| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
444 PKPNPN-----NPNPNPNPNPNPNPN-----EPLNPNPNPNPNPNPNPNPNPNPN 492
554 PAPTTPKKAAPTTPKEP-APTTPKETAPTTPKLLPTTPEKLAAPTTPKEKAPATTPPELAP 612
| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
493 PS-----NPNPNPNPNPNPNPN-----PSNPK-----PSNPNP--P 523
613 TTTPEEPPTTTPPEAPATTPKAAAPNTPKPE-APTTPKEAPATTPKKEAPATTPKETAPTTP 671
524 SNPN-----PLNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 564

```

[illegible]



```

: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 2626
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1185 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-041-886-23

```

```

Query Match      5.88; Score 425.5; DB 4; Length 1185;
Best Local Similarity 21.88; Pred. No. 1.5e-19;
Matches 240; Conservative 119; Mismatches 437; Indels 303; Gaps 52;

QY 88 SQTIKSTTKRSPPKPNKKKKTKVIESEITEHSEVSENOESSSSSSSSSTTWIKKS 147
DB 9 SMSNRSGKKKAPGR-----BELNRGRASPGGVSTSSDG-----KAERK 50
QY 148 KNSAANRELAKLKVKNKKRKKPKPPVVDGASGLDNGDFKVTTPDTSTQHNK 207
DB 51 RQTA-----KARVEASTPK--VNKGRS-----ELISEESE 82
QY 208 VTSPTKITTAKPIPNRPSLPNSDTSKETSLTVNKEETVETKETTNTKOTSDGKEKT 267
DB 83 ETNAPKTKTEQELRPOSQSDLDLDGRSL--NDGSSPDRDQONRST-----P 133
QY 268 SAKETOSIEKTSADLAPTSVLAAPKPKAETTKGPAITPKRPPTTP--KEPATTP 325
DB 134 SIYSPGSEVDS-----DSSGSLSGPA--RPIHPRLPPSPQRPDSTP 176
QY 326 KEPTPTTIKSAPTTPKREAPRTTKSAPTTPKEAPRTTKREAPRTTPKEAPRT 385
DB 177 ROPASAP-----EPHPSVPTGYHAPMER--PTSRMGOAR--PARPRPHQLYGGT 224
QY 386 --TTKSAPTTPK-----EAPRTTPKRAAPRTTPKEAPRTTPKEAPRTTP 427
DB 225 GGVLSGPMGRKGGGAASVGRNGGKONRPRTTPISVSSSGASGAPRTK--PTTPV 280
QY 428 EPAPTTKEPAP-----TPKEAPRTA--PKKRAPTTPKEAPRTTPKEAPRTTP 480
DB 281 GGGNLBPARRPARNPTPLRRLRPLRNANASAP-----PGLGAORPLGHL--PSRYA 334
QY 481 PKE-----PAPTTTSAPRTTPKEP--APTTTSAPRTTP--KEPSPTTKREAPRTTPKEPA 531
DB 335 MCGGMGLRPGP---EKGPPLASPHSLRPASSAPAPARPRPYSSSSSSSAAASSSSS 391
QY 532 PTTPKKAAPTTPKEAPRTTPKE--PAPTTTKKRAPTAPKEAPRTTPKTAATTPK 587
DB 392 SSSSASPPRAS--QALDSYHSPRPSTL-----SVSNOP--PKYTOPSLPQAVMSQ 440
QY 588 --PTTPE--KLAPTTKEKAPRTTPEELAPRTTPEEPRT----- 621
DB 441 GPRPPRYGRLNLSNHNRRPFRSTGASQASTAPRVSTNNHNNQOOOQOOOQOOHNG 500
QY 622 --TPREAP-----TPKKAARN-----TPKEAPRTTPKEP-----APTPKE 657
DB 501 NSGPRPGAPRHLLEGSSHNAPYAMSPSLGLRYPRRPRAHLRPHNSQVSYSGAPNG 560
QY 658 PAPTPKREAPRTTPKGAAPTTLKEAPRTTPKKAAPKELAP-----TTTKEPTSTSDKAPRT 714
DB 561 PYSVSSSSNSSSTSGSYPCS--HPSPOGPOGAPYRPPRVVTVTSSATLSTVIAVTAS 618
QY 715 TPFGIAPPTTPKEAPRTTPKGAAPTTLKEAPRTTPKKAAPKELAPRTTPKAPTS 774
DB 619 SPAGYKTASPPGPPYGGKAPSPGAYKTATPRPYKPGSP-----PSPRGTGTPRGYNG--- 670

```

```

QY 775 TTSKPAAPTPKETAAPTTPKREAPRTTPKKAAPRTTPETPPTTSEVSTPTTKPTTHKS 834
DB 671 -TSPAGGTTPKPSPTV--GGPLPPAGPS--GLPSLPPPAAPASBPPLS--ATOIKOE 724
QY 835 PDE--STPELSAEPKALENS--PKERGVTTPKPAATKREMTTAKDKTETEDLTTP 890
DB 725 PAEEYETPE--SPVPPARSPSPPKVVDVSHASQSRKFKHLDRGPNSCARBDLTFVP 781
QY 891 ETTAAPKMKETATTTTEKT--TESKIYATTTQVSTTTQDTPPTPKITTLKTTTLAPVYT 949
DB 782 --LEGSKLAKRADLVEKYRREAQORAR----- 807
QY 950 TKKTTITTEIMNKKPEELAKPRODATNSKATTPKPOKPTKAPKKTSTKKRTMPRVKPK 1009
DB 808 -----EKKERERERERERERERERERERERERERERERERERERERERERERER 848
QY 1010 TTPTRKMTSTMPELNPTSRIAEAMLOTTPRNPOTPKSKLVEVNPKSEDAAGAGETPHM 1069
DB 849 LGPVPHR-----PPEERGSANA-----IYPPYLGPDTPALRT--LSEYA----- 885
QY 1070 LRPVHFMPEVTPDMYLP 1088
DB 886 --REHVMSPGNRNHPFYVP 902

```

```

RESULT 12
US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, Pierre
; APPLICANT: LATHE, Marie-Paule
; APPLICANT: HAREVENT, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITILE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22133-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021

```

```

: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1867 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FEATURE:
:   NAME/KEY: Peptide
:   LOCATION: 128..1727
:   OTHER INFORMATION: /note= "The amino acids spanning
:   OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
:   OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
:   OTHER INFORMATION: repeats varies from 1 to 40."
: FEATURE:
:   NAME/KEY: Peptide
:   LOCATION: 134
:   OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
:   OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
:   OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
: FEATURE:
:   NAME/KEY: Peptide
:   LOCATION: 144
:   OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
:   OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
:   OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
: FEATURE:
:   NAME/KEY: Peptide
:   LOCATION: 147
:   OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
:   OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
:   OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
: FEATURE:
:   NAME/KEY: Peptide
:   LOCATION: 1..21
:   OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
:   OTHER INFORMATION: 21 amino acid precursor sequence."
: US-08-479-537A-5

```

```

: 420 SAPDXRPGSTAPXAHGVTSA-----APDXRPGSTAPXAHGVTSA-----PGST 471
: 630 TPRA-----AAPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAP 664
: 472 APXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGST 530
: 685 TTP-----KKPAPKELAPPTTKEPTSTSDKAP--TTPKGAAPTTPKEPAPPTTP 732
: 531 TAPXAHGVTSAAPDXRPGSTAP--XAHGVTSAAPDXRPGSTAPXAHGVTSA-----APDXR 585
: 733 KEPAPTTPKPG-----TAPTTTKEPAPPTTP-----KKPAPKELAPPTTTPKEPTSTSD 778
: 586 PXPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAP--XAHGVTSAAPDX 644
: 779 KPAP--TTPK-----ETAPTTTKEPAPPTTP-----KKPAPTTTETPTTTPPTTSEVSTP 822
: 645 RPXPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAP--TAPXAHGVTSA 702
: 823 TTPKEPT-----HKSPDESTPELSAETTPKALENSPREKPGVPTTKTPAA----- 868
: 703 DXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSA 762
: 869 -TKPEMTTAKDK--TTERDLRTTPETTTAAPTMTKETATTEKTESKITATTTQVTS 924
: 763 DXRPGSTAPXAHGVTSAAPDXRPGST--APXAHGVTSAAPDXRPGSTAPXAHGVTSA 820
: 925 TTQDTPPKITTLTKTTLAKVTTTKITITTEIMKKPEETAKPKRATNSKATTPKPK 984
: 821 APDXRPGXG-----STAPXAHGVT-----SAPXRPXPGSTAPXAHGVTSAAPDXRPGX 866
: 985 KPTAKPKPTSTKKRKTTPRYRKKTTPTPRKMTSTPELNPTRIAEAMLQTTTRPNQT 1044
: 869 -GSTAPXAHGVTSAAPDXRPG--XPGSTAPXAHGVTSA--PDXRPG-----PGST 911
: 1045 -PNSKLVENPKSEDAAGABEETHMLRPVPEVTPPMDIYLPVYVNOGIIINPLMSD 1103
: 912 APXAHGVTSAAPDXRPGSTAPXAH-----GVTSAAPDXRPGSTAP--PXAHG 957
: 1104 ETNLCNGKPVVDGLTTP--LNGGLVAFRGHYFWMLSPPSPSPARRITFVWGIPSPID 1158
: 958 VTSAPDXRPGSTAPXAHGVTSA-----PDXRPGSTAPXAHGVTSAAPD 1003

```

RESULT 13  
 US-09-083-116-5  
 : Sequence 5, Application US/09083116  
 : Patent No. 6203795  
 : GENERAL INFORMATION:  
 : APPLICANT: CHAMRON, Pierre  
 : APPLICANT: KIENVY, Marie-Paule  
 : APPLICANT: LATHE, Richard  
 : APPLICANT: HAREUVENT, Mara  
 : TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
 : TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
 : NUMBER OF SEQUENCES: 5  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
 : STREET: P. O. Box 1404  
 : CITY: Alexandria  
 : STATE: Virginia  
 : COUNTRY: United States  
 : ZIP: 22313-1404  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/083,116  
 : FILING DATE:  
 : CLASSIFICATION:  
 : PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Rodin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-5

Query Match 5.8%; Score 424.5; DB 4; Length 1867;
Best Local Similarity 23.7%; Pred. No. 2.9e-19;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

```

```

379 TKE--PAPTTT-----KSAPTTKREPAPTTK-----KPAPTTKREPAPTTKREPPTT 425
Db 181 APDXRPXPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 425
Qy 426 PKE--PAPTTKREP-----APTTKREPAPTTK-----KPAPTTKREPAPTTKREPAPTT 472
Db 241 APDXRPXPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 300
Qy 473 TKESPPTTKREPAPTT--TKSAPTTTKREPAPTT-----TKSAPTTKREPSPPTTK-----E 521
Db 301 APDXRP--XPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 359
Qy 522 PAPTTKREPAPTTK-----KPAPTTKREPAPTTK-----EPAPTTTKKAPTTKREPAPTT 573
Db 360 SAPDXRPXPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 419
Qy 574 TPKEPTAP-----TPPKLTPPTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTT 629
Db 420 SAPDXRPXPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 471
Qy 630 TPKA-----AAPNTPKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTT 684
Db 472 APXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPX 530
Qy 685 TTP-----KKAPKELAPTTTKEPTSTSDKAP--TTPKGTAPTTKREPAPTTT 732
Db 531 TAPXAHGVTSAPODXRPXGSTAP--XAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAP 585
Qy 733 KEAPPTPKG-----TAPPTLKREPAPTTT-----KKAPKELAPTTTKEPTSTSD 778
Db 586 PAXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAP--XAHGVTSAPODX 644
Qy 779 KPAP--TTPK-----ETAPTTKREPAPTTT-----KKAPPTTKEPTSTSEVSTP 822
Db 645 RXPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAP--TAPXAHGVTSAPO 702
Qy 823 TTTKEPTT-----IHNSPDESTPELSAEPPTPKLENSPKKEPGVTTKTPAA----- 868
Db 703 DXRPXPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPO 762
Qy 869 -TKREMTTAAKDK--TTERDLRTPEPTTAAPKMTKETAATTEKSTTATTTQVTS 924
Db 763 DXRPXPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPO 820
Qy 925 TTTQDTTPPKITLTTTLARVTTTAKTITTTTETIMNKPEETAPKDRATNSKATTPKQ 984
Db 821 APDXRPXG-----STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTSAPODXRPX 868
Qy 985 KPTKAPKKRPTSTKRPKTPRVKRPKKTPTTPPKMKSTMPDELNPTSRIDEMLQTTTRPNOT 1044
Db 869 -GSTAPXAHGVTSAPODXRP--XPGSTAPXAHGVTSA--PDKRPX-----PGST 911
Qy 1045 -PNSKLEVNPKSDDAGAGETPHMLLRPHVPEVTPMDVLPVPRNPGIITINPLSD 1103
Db 912 APXAHGVTSAPODXRPXGSTAPXAH-----GVTSAPODXRPXGSTAP--PAXHG 957
Qy 1104 ETNICKGKPVVDGLTT--LRNGTLVAFRGHYFMMLSPFEPSPARITIEWGISTPDPID 1158
Db 958 VTSAPDXRPXGSTAPXAHGVTSA-----PDXRPXGSTAPXAHGVTSAPODXRPXGSTAP 1003

RESULT 14
US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHEY, Richard
; APPLICANT: HAREDEVINT, Maria
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

```

NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BURNS, DOANE, SWECKER & MATTHIS, L.L.P.  
 STREET: P.O. Box 1404  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/479,537A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FR 90/13101  
 FILING DATE: 23-OCT-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: NO PCT/FR91/00835  
 FILING DATE: 23-OCT-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/039,320  
 FILING DATE: 04-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/403,576  
 FILING DATE: 14-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Teskin, Robin L.  
 REGISTRATION NUMBER: 35,030  
 REFERENCE/DOCKET NUMBER: 017753-025  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2035 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 128..1899  
 OTHER INFORMATION: /note= "The amino acids spanning  
 OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat  
 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
 OTHER INFORMATION: repeats varies from 1 to 40."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 134  
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
 OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,  
 OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."  
 OTHER INFORMATION:  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 144  
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA  
 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
 OTHER INFORMATION:  
 NAME/KEY: Peptide  
 LOCATION: 147  
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 1..21  
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
 OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2  
 Query Match 5.8%; Score 424.5; DB 2; Length 2035;  
 Best Local Similarity 23.7%; Pred. No. 3,2e-19;  
 Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;  
 236 TSLTVNKEETVEYKETTNNKOTSDGKEKTSKKEQSIKTSKARD-LAPTSKVLAKPT 294  
 16 TVLTV-----VYSGHASTPGGEKETSATQSSVSPSSTEKNAVEMTSSVLSHS 65  
 295 P-KAETTTKGP--ALTTPKEP-----PTTPKEPASTTPKEPTTTKSA 337  
 66 PGSSGTTGGQDYLAPATEPAGSAAATWGQDVTSVPTRPALOSTTPPAHDVY---SAP 122  
 338 TTPKEPAPTT-----KSAPTTPKEP-----APTTPKEPAPTTPKEPAPT 378  
 123 --DNKPAFGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 180  
 379 TKE--PAPTT-----KSAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPTT 425  
 181 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 240  
 426 PKE--PAPTTKBP-----APTTPKEPAPTAPK---KPAPTTPKEPAPTTPKEPAPT 472  
 241 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 300  
 473 TKEPSTTPKEPAPT--TKSAPTTPKEPAPT-----TKSAPTTPKEPSPPTTK----E 521  
 301 APDXRP--XPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 359  
 522 PAPTTKEPAPTTPK---KPAPTTPKEPAPTTPK---EPAPTTPKAPAPAPKEPAPT 573  
 360 SAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 419  
 574 TPKEAP-----TPPKLITTPTEKLAPTTPKEPAPTTPKEAPTPPEPTTPPEPAPT 629  
 420 SAPDXRXPXGSTAPXAHGVT---APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 471  
 630 TPKA---AAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKEAPTPPEPTTPPEPAPT 684  
 472 APXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 530  
 685 TTP-----KKPAPELAPTTTKEPTSTSDKAP--TPPKTAPTPTPKEPAPTTP 732  
 531 TAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDXRXPXGSTAPXAHGVT---APDXR 585  
 733 KEPAPTTPKGS---TAPTTLKEPAPTTP-----KKPAPELAPTTTKEPTSTSD 778  
 586 PXPOSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDX 644  
 779 KPAP--TTP---ETAPTTTPKEPAPTTP-----KKPAPTTPPTTPPTSEVTP 822  
 645 RXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXG--TAPXAHGVTAP 702  
 823 TTTTEPTT-----HKSPDESTPELSAPETPKALENSEKEGVPTTKTPAA----- 868  
 703 DXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAP 762  
 869 -TKPEMTTAKDK---TTERDLTTPETTTAAPKMKETATTEKTESITATTQYNS 924  
 763 DXRXPXGSTAPXAHGVTAPDXRXPXGST--APXAHGVTAPDXRXPXGSTAPXAHGVT 820  
 925 TTTQDTTFKTLITLAKVTTTKKTIITTIMKPEETAPKPKDRAVNSKATTPKPQ 984  
 821 APDXRXPXG-----STAPXAHGVT-----SAPXRPXGSTAPXAHGVTAPDXRXP 868  
 985 KPTKAPKPTSTKKKPTTPRVKPKTTPTPKKMTSTPELNPISRIAEMLQTTTRNOT 1044  
 869 -GSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTSA--PDXRXP-----PGST 911  
 1045 -PNSKLVENVKSEDPAGAGETTPHMLLRHVHVEVTPDMDVLPVAVPNOGIIINPLMSD 1103



Db 912 APXAHGVTAPDXRXPXGSTAPXAH-----GVTAPDXRXPXGSTA---PXAHG 957  
QY 1104 ETNINCGRPVDGLTT-LRNGTLVAFRGHYFWMLSPSPSPARRITEVWGISPID 1158  
Db 958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTSA 1003

RESULT 15  
US-09-083-116-2  
; Sequence 2, Application US/09083116  
; Patent No. 6203795  
; GENERAL INFORMATION:  
; APPLICANT: CHAMON, Pierie  
; APPLICANT: KIENY, Marie-Paule  
; APPLICANT: LATHE, Richard  
; APPLICANT: HAREUVENI, Mara  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,116  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/479,537  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/FR91/00835  
; FILING DATE: 23-OCT-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/039,320  
; FILING DATE: 04-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/403,576  
; FILING DATE: 14-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Teskin, Robin L.  
; REGISTRATION NUMBER: 35,030  
; REFERENCE/DOCKET NUMBER: 01753-025  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-6620  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2035 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 128..1899  
; OTHER INFORMATION: /note= "The amino acids spanning  
; OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat  
; OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
; OTHER INFORMATION: repeats varies from 1 to 40."  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 134  
; OTHER INFORMATION: /note= "amino acid 134 is X1 - Xaa  
; OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro - CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala - GCT, GCC, GCA, or GCG."  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 144  
; OTHER INFORMATION: /note= "Amino acid 144 is Y - Xaa  
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr - ACT, ACC,  
; OTHER INFORMATION: or ACG; and Asn - AAT or AAC."  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 147  
; OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa  
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, CCC,  
; OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..21  
; OTHER INFORMATION: /note= "amino acids 1 to 21 are a  
; OTHER INFORMATION: 21 amino acid precursor sequence."  
US-09-083-116-2

Query Match 5.8%; Score 424.5; DB 4; Length 2035;  
Best Local Similarity 23.7%; Pred. No. 3.2e-19;  
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 236 TSLTVNKETVETMETTTKQSTDKKETSKEKTSKAKD-LAPTSKVLAKPT 294  
Db 16 TVLTV-----VSGSHASSTPGKEKETSATORSVSSPEKNAVMTSSVLSHS 65  
QY 295 P-KAETTTKGP--ALTTPKP-----PTTPKEPASTPKKEPTTTIKSAP 337  
Db 66 PGSSSTTGODVYLAATEPAGSAAATNGODVTSVVTTPALGSTTPPAHDVT---SAP 122  
QY 338 TTPKEPAPTT-----KSAPTTKEP-----APTTPKEPAPTTKEPAPTT 378  
Db 123 --DKRPAAGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 180  
QY 379 TKE--PAPTT-----KSAPTTKEPAPTTK-----KAPTTPKEPAPTTKEPPTT 425  
Db 181 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 240  
QY 426 PKE--PAPTTKEP-----APTTPKEPAPTAPE--KPAPTTPKEPAPTTKEPAPTT 472  
Db 241 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 300  
QY 473 TKEPSPTTKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTKEPSPPTTK----E 521  
Db 301 APDXRP--XPSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 359  
QY 522 PAPTTPKEPAPTTK-----KPAPTTPKEPAPTTK-----EPATTTKKAPAPTAPEAP 573  
Db 360 SAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 419  
QY 574 TPKEAP--TTPKRLPTTPPEKLAPTTPPEKAPPTPELAAPTTPPEEPPTTPPEAPPT 629  
Db 420 SAPDXRXPXGSTAPXAHGVT-----APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAP 471  
QY 630 TPKA-----AAPNTPKREAPPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTT 684  
Db 472 APXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXP--XPSTAPXAHGVTAPDXRXPXGS 530  
QY 685 TTP-----KKPAPELAAPTTPKEPTSTTSOKPAP--TTPKGTAPTTTKEPAPTTPT 732  
Db 531 TAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDXRXPXGSTAPXAHGVT-----APDXR 585  
QY 733 KEPAPTTPG-----TAPTTLKEPAPTT-----KKPAPELAAPTTPKTGPTSTSD 778  
Db 586 PXPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDX 644  
QY 779 KPAP--TTPK-----EPATTPKEPAPTT-----KKAPPTTPETPTTSEVSTP 822  
Db 645 RXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGS--TAPXAHGVTAP 702

```

QY 823 TTtKEPTT-----IHKSPDESTPELSAEPTEKALENSPEKPEGVPTTKTPAA----- 868
Db 703 DXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PD 762
QY 869 -TRPEMTTAKDK---TTERDLRTTPETTAARPKMTKETATTEKITESKITATTOYTS 924
Db 763 DXRPXPGSTAPXAHGVTSA PDXRPXPGST--APXAHGVTSA PDXRPXPGSTAPXAHGVT 820
QY 925 TTQODTTPFKITTLTKTTLAPKVTTTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKPQ 984
Db 821 APDXRPXPG-----STAPXAHGVT-----SAPDXRPXPGSTAPXAHGVTSA PDXRPX- 868
QY 985 KPTKAPKPKPTSTKKPKTMPRVKPKKTTPTPRKMTSTMPBELNPTSRIAEAMLQTTTRPNQT 1044
Db 869 -GSTAPXAHGVTSA PDXRP--XPGSTAPXAHGVTSA--PDXRPX-----PGST 911
QY 1045 -PNSKLVEVNFKSEDAGABGETPHMLLRPHVFMPEVTPDMDYLRVYVNOGIIINPMLSD 1103
Db 912 APXAHGVTSA PDXRPXPGSTAPXAH-----GVTSA PDXRPXPGSTA---PXAHG 957
QY 1104 ETNIONGKPVVDGLTT-LRNGTLVAFRGHYFWMLSFSPSPPARRITEVWGIPSPID 1158
Db 958 VTSAPDXRPXPGSTAPXAHGVTSA-----PDXRPXPGSTAPXAHGVTSA PD 1003

```

Search completed: April 26, 2002, 16:26:07  
 Job time: 507 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:20 ; Search time 114.61 Seconds  
(without alignments)  
904.577 Million cell updates/sec

Title: AA7  
Perfect score: 7323  
Sequence: 1 MAMKTLPIYLLLSVFVIO.....ARAITRSQGLSKWVWVNC 1361

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR68:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1088.5	14.9	3020	2 A43932	mucin 2 precursor,
2	950	13.0	1664	2 T18262	S-layer protein -
3	867	11.8	1489	2 T31108	cyst germination s
4	835	11.4	1274	2 T16251	hypothetical prote
5	828.5	11.3	2187	2 T30826	nascent polypeptid
6	800	10.9	1188	2 S49915	extensin-like prot
7	795.5	10.9	1367	2 S48478	glucan 1,4-alpha-g
8	704.5	9.6	1151	2 T18535	high molecular mas
9	674	9.2	1229	2 T25697	hypothetical prote
10	672	9.2	1344	2 A35175	mucin 1 precursor,
11	670	9.1	7962	2 T38346	elastic titlin - hu
12	668	9.1	3507	2 T34513	hypothetical prote
13	663	9.1	3570	2 T45025	mucin MUC5B, trach
14	646	8.8	990	2 I51618	nucleolar phosphop
15	633	8.6	489	2 T11622	extensin class 1 p
16	632	8.6	761	2 C84672	hypothetical prote
17	632	8.5	971	2 T19431	hypothetical prote
18	626	8.5	6642	2 T29757	protein UNC-69 - C
19	622.5	8.3	839	2 T75518	hypothetical prote
20	607.5	8.3	801	2 T29018	hypothetical prote
21	607.5	8.3	924	2 S27923	gene lpf protein -
22	605	8.3	379	2 S50125	larval glue protei
23	568	7.8	2232	2 T34434	hypothetical prote
24	566.5	7.7	875	2 S23760	polyphenolic adhes
25	565	7.7	1630	2 A53577	ascites stialoglyco
26	559.5	7.6	350	2 S22456	hydroxyproline-ric
27	556.5	7.6	856	2 T15543	hypothetical prote
28	551	7.5	620	2 S06733	hydroxyproline-ric
29	545	7.4	873	2 A47283	calphostin - fruit

30	542	7.4	369	2 S20500	hydroxyproline-ric
31	538.5	7.4	416	2 T00465	extensin precursor
32	535	7.3	756	2 T27642	hypothetical prote
33	533	7.3	1087	1 QEMSH	neurofilament trip
34	530.5	7.2	1162	2 T40557	exo-alpha-stallidas
35	530	7.2	865	2 A47282	calcium-binding pr
36	528	7.2	3534	2 T42567	tegment protein 2
37	527.5	7.2	1459	2 T32271	hypothetical prote
38	522	7.1	328	2 T00985	hydroxyproline-ric
39	519	7.1	813	2 S70795	vsaA protein precu
40	518	7.1	866	2 T45462	membrane glycoprot
41	518	7.1	1072	2 A37221	neurofilament trip
42	514	7.0	1611	2 T38236	hypothetical prote
43	512.5	7.0	867	2 T45463	membrane glycoprot
44	510.5	7.0	1832	2 T31113	mucin-like glycopr
45	509.5	7.0	662	2 A45155	mucin FTM-C.1 - Af

## ALIGNMENTS

RESULT 1  
A43932  
mucin 2 precursor, intestinal - human (fragments)  
N:Alternate names: mucin SMUC-41  
C:Species: Homo sapiens (man)  
C:Date: 10-Mar-1993 #sequence, revision 12-Apr-1996 #text change 05-Nov-1999  
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t  
A:Reference number: A49963; MUID:94132002  
A:Accession: A49963  
A:Molecule type: mRNA  
A:Residues: 1-639 <GU1>  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
A:Reference number: A45106; MUID:93016075  
A:Accession: A45106  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 626-1895 <GU2>  
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396  
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A:Accession: B45106  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 2037-3020 <GU3>  
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398  
A:Experimental source: colon  
A:Note: sequence extracted from NCBI backbone (NCBIP:116988)  
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,  
J. Clin. Invest. 86, 1005-1013, 1991  
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym  
A:Reference number: A43932; MUID:91358717  
A:Accession: A43932  
A:Molecule type: DNA  
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864  
A:Note: sequence inconsistent with the nucleotide translation  
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl  
A:Reference number: A33532; MUID:9197956  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GU4>  
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874  
A:Experimental source: intestine  
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.









QY 647 PTPKKEPAPPTPKETAPPTPKGTAPPTLKE---PA-----PTTPKK---DAPKELAPPTT 695  
 Db 918 PAAVSSP-PTTPKSSPPPVVSSPPVVKSSPPAPVSSPPAPVSSPPAPVNLPEV 976  
 QY 696 KE---PTSTTSKDPAPPTPKGTAPPTP-KEPAPPTPKETAPPTPKGTAPPTLKEPAPPTT 751  
 Db 977 KSSPPPTPVSSPPA---PKSSPPAPVMSPPPEVKSSPPAPVSSPPPVKSSPPAP 1033  
 QY 752 KKPAPKELAPPTTKGPTSTSDKAPPT-TPKAPPTTPKAPPTPKKAPPTPEPTPTT 810  
 Db 1034 VSSPP---PVPKSSPPAPVSSPPPVKSSPPAPVSSPPPVKSSPPAPVSS-SPPPP 1088  
 QY 811 TSEVSPPTTKEPTTHKSPDESSTPELSAPPTKALENSPKKEGVPPTTKTAPATKPEMTT 870  
 Db 1089 VKSPPPAPVSSPPPKSPPPAPVSSPPAPVSSPPAPVSSPPAPVSSPPAPV 1141  
 QY 871 TAPKKTTERDLRTPTT 887  
 Db 1142 PAPPKKEQSLPPAPES 1158  
 RESULT 7  
 S48478  
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019c  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 12-Nov-1999  
 C:Accession: S48478; A26877; B26877; S27281; J06123  
 R:Rowley, K.  
 submitted to the EMBL Data Library, October 1994  
 A:Reference number: S48478  
 A:Accession: S48478  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <ROW>  
 A:Cross-references: GB:Z47047; EMBL:Z38061; NID:9603997; PID:9763364; GSPDB:GN00009; MIF  
 R:Yamashita, I.; Nakamura, M.; Fukui, S.  
 J. Bacteriol. 169, 2142-2149, 1987  
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.  
 A:Reference number: A91831; MUID:87194600  
 A:Accession: A26877  
 A:Molecule type: DNA  
 A:Residues: 1-242 <YAM>  
 A:Cross-references: EMBL:M16164; NID:9172522; PIDN:AAA35014.1; PID:9172525  
 A:Accession: B26877  
 A:Molecule type: DNA  
 A:Residues: 762-1331 <YA2>  
 A:Cross-references: EMBL:M16165; NID:9172523; PIDN:AAA35015.1; PID:9172526  
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
 FEBS Lett. 239, 179-184, 1988  
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar  
 A:Reference number: S27281; MUID:89031230  
 A:Accession: S27281  
 A:Molecule type: DNA  
 A:Residues: 1-31 <PAR>  
 A:Cross-references: EMBL:X13857; NID:94551; PIDN:CAA32069.1; PID:94552  
 R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
 A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudohy  
 A:Reference number: J06123; MUID:96523237  
 A:Accession: J06123  
 A:Molecule type: DNA  
 A:Status: nucleic acid sequence not shown  
 A:Residues: 1-1367 <LAM>  
 A:Cross-references: GB:U30626; NID:91304386; PIDN:AAC49609.1; PID:91304387  
 C:Genetics:  
 A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
 A:Cross-references: MIPS:YIR019c; SGD:S0001458  
 A:Map position: 9R  
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
 F:1350-1366/Domain: transmembrane #status predicted <TM1>  
 F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 10.9%; Score 795.5; DB 1; Length 1367;  
 Best Local Similarity 27.4%; Pred. No. 2,3e-26;  
 Matches 318; Conservative 111; Mismatches 525; Indels 207; Gaps 51;  
 85 DCDACCKKDKD---CCDYSESFCAEVHNPTSPSSKAPPPSGASQNTKTKKSPKPN 141  
 Db 174 DLSTGCNNYDNOGHSQDFDFGFWNIDCNDNGGKSTTSSSES- -STTSS----- 226  
 QY 142 KKTAKVIESEETEVKDNKNRTKKKPKPPVVDAGSLDNGDKVTPDSTIQHN 201  
 Db 227 -----TSSSTTSSSTSSST----- -TTSSSTSSSTSS 253  
 QY 202 KVTSPKPI-TTAKPINSPLPNSDTSKETSLSVKNETVEKETTTKQTSDOCKE 260  
 Db 254 SSTTAPAPPTTSCTKKPKPPPTTSCTKKPPPHHDTPCKKTKTTSK-TCT---KK 309  
 QY 261 TTSKETSIEKTSKADLPTSKVLAKPTPKAETT--TKGPALTTPKE-----PTPTT 312  
 Db 310 TTPVPPTPS- -STTESSAPV-----PTSSSTTESSAPVSSSTTESSAPVPTSSS 362  
 QY 313 KEPASTTPKEPTPTTKSAP---TPKEPAPTTKSAPTTPKEPAPTTKEPAPTTKE 368  
 Db 363 TTESSAPVNTSSTTESSAPVNTSSTTESSAPVPTSSSTTESSAPVNTS- -TTESS 418  
 QY 369 PAPTTKEPAPTTKSAPTTPKEPAPTTKPKAPPTPKKAPPTPKKPTT- -TPKEPA 425  
 Db 419 SAPVTS- -STTESSAPVNT- -STTESSAPVNTSSTTESSAPVPTSSSTTESSSA 471  
 QY 426 PTPKEPAPTTKEPAPTTKPKAPTTKEPAP--- -TPKEPAPTTKESPTTPKEPAP 481  
 Db 472 PVT- -STTESSAP- -VPTSSSTTESSAPVNTSSTTESSAPVPTPS- -STTESSAP 526  
 QY 482 TTKSAPTTKKEPAP--- -TTKSAPTTKPEPPTTKKEPAP--- -TPKEPAP-TPK 532  
 Db 527 APPTSSSTTESSAPVNTSSTTESSAPVPTSSSTTESSSTPTVSSSTTESSAPVPTSS 586  
 QY 533 PAPTTKEPAPTTKKEPAPTTKPKAPTTKPKAPTTKPKAPTTKPKAPTTKPKAPTT 592  
 Db 587 STTESSAPVPT- -PSSSTTESSAPVPTSSSTTESSAPVPT- -STTESSAP- 637  
 QY 593 TPKEPAPTTPELAPTT-TPKEPPTTPPEEPAP--- -TPKAAAPNTPKKEPAPTTKEPA 646  
 Db 638 VPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPT- -SSTTESSA 693  
 QY 647 PTPKKEPAPTTKEPAPTT-TPKGTAPTTKEPAPTTKPKAPKELAPTTKEPTSTSDK 705  
 Db 694 PVT- -SSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTPS- -SSTTESS 748  
 QY 706 PAPTTPKGTAPTTKEPAP--- -TPKEPAPTT-TPKGTAPTTKEPAPTTKPKAPKELAP 761  
 Db 749 SAPVTSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPV 808  
 QY 762 TTKGPTSTSDKAPPTPKETAPTTKEPAP--- -TPKKEPAPTTPEPTTSEVST 816  
 Db 809 VPTSSSNNTSSAPSTPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPT 868  
 QY 817 PT-----TKPEPTTHKSPDESSTPELSAEPPTKALENSPKKEGVPPTTKTAPATKPEMTT 871  
 Db 869 PSSSSNNTSSAPSI- -PSSSTTESSSTGT- -TVPSSSKVPSQETSVSTTEVTV 924  
 QY 872 AKDKTTERDLRTPTTETTTAARKMTKEATTEKTESKIRAT--TQVNSTTQODTTPK 929  
 Db 925 TKTTTSTTTPSTTTTITTVOCSTGNSAGETTSKSPKVTITVPTTITTVSVTSSITTTIT 964  
 QY 930 ITTLKTTTLAPKVT- -TKKTTTT-ELMNKDEETAKPKDRATNSKATPKPKPK- -981  
 Db 985 TTVCSGNTNAGETTSKSPKITTITVPCSTSEIA-----SESTTSTPTTPTTV 1036  
 QY 982 -----TKAPKETSCKKKTMDRVKPKTTPTPRMTSTMPELNPTSRIAEA 1028  
 Db 1037 VSTTVVTVETSTSTKPGGETTTTFVTKNIPVTLLTIAPPT- -SVTVVTVNFTPTT-ITTT 1093



Query 1029 MLOTTTRPNOTPNSKLVENPKSEDAGAGETPHMLRPHVEMPEVTPMDLPRVPO 1088  
 1094 VCSIGF-----NSAGETSGSPKVTVTTPCSTGTGETTATTLVTAVT 1140  
 QY 1089 GIIINMLSDENICNGKPPDGLITLNGTLVAFRGHYEMLSFSPSPSARITEWGI 1148  
 Db 1141 TTVVTSSSTGTNSA-GKTTTGYTKSVPTT-----YVTTLPASAPVPAFN-----AV 1188  
 QY 1149 PSPIDVTTRC-----NCEGKT 1165  
 Db 1189 PTTTIT--TECSAATNAGET 1207

RESULT 8  
 T18535  
 high molecular mass nuclear antigen - chicken (fragment)  
 C:Species: Gallus gallus (chicken)  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18535  
 R:Shimada, K.; Harata, M.; Mizuno, S.  
 J. Cell Sci. 110, 3031-3041, 1997  
 A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, hNNA, of chick  
 A:Reference number: Z18955; M0ID:9803440  
 A:Accession: T18535  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1151 <SH1>  
 A:Cross-references: EMBL:D88440; NID:d117138; PID:d1025045; P1DN:BA24137.1

Query Match 9.6%; Score 704.5; DB 2; Length 1151;  
 Best Local Similarity 25.7%; Pred. No. 1.2e-22;  
 Matches 304; Conservative 104; Mismatches 445; Indels 331; Gaps 59;

QY 105 AEVHNPTSPSSKK-----APPSG-----ASOTIKSTTKRSPKP 140  
 Db 39 AEHL-PAAPQPPKWPVIGAPPPEPTPTPSKPTDGADAAAPKASELTSPPASPSPP 97  
 QY 141 NKKTKKVVIESEITEVKDKKKRTKKPKP-----PVVDEAGSGDNGDFKVTTP 193  
 Db 98 DGPKAPSGAGEAP-----AGTPPSGCPAGTPPSGAGAPKCGDTAOPS 143  
 QY 194 DYSTTQHNKVS--PKITTAKPINRPSLPPNSDTSKETSLTVNKETVETKETTTNK 251  
 Db 144 GTSAGADGKPAQDVPKATTA-ATEAR-----ASAAPVYKATATATVTAASQS 194  
 QY 252 --QTSIDGKEKTTSAKETGSEKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPT 309  
 Db 195 APKAAADAAAVTAA--SOSAPKATV-EVKPAAAVAAKAKAVTAAAAAAPKATAEAKPAP 250  
 QY 310 -TPPKREPASTPKP-----TPTIKSNAPTPPKREPAPT-----TKSAPTPPKP 353  
 Db 251 VISPPTPCSSAEKPLTAASPTASKA--TAEAKVPATASLMAKTATAEKAPSPSVK 308  
 QY 354 APPTTK-----EPAPTTPKEPAPTTPPKREPAPTTPKSAPTT 388  
 Db 309 ATDTKAVATATAPKAGPDYKPAVAVCAEAKPAPPPQOPKAAAAAAGTGLKATPAT 368  
 QY 389 -----PREPA-----PTTPKKRAPPTTPKEPAPTTPKEPTT-----PKRAPTTPKEPAP 433  
 Db 369 PHGSPRANSHVTVPNVPRAAAATVP--TAGAVPKASGTTPPAAAPQOPVPV--KAAP 423  
 QY 434 TPKEPAPTAPKAPAPTPPKREPAPTTPKEP-----APTTPKPSPTTPKEPAPTTPKSAPTT 490  
 Db 424 VTPSPQOAVPRAATAAA-----APVTPOQVYTAATTTNATPPPOPIPKAATTTTATPVT 479  
 QY 491 TKEPAP-----TTTKSAPPTTPKEPSPTTPKEP-----APTTPKEPAPTTPKRAPPTTPKEPAP 543  
 Db 480 PQOPIPKAGIDAAPPAVPKAPSDGRAATPGVNAATDPKPPPTQSVSPSAVTEPRQP 539  
 QY 544 -----TPKEPAPT--TTKKRAPTAPK-----EPAPTTPK--ETAPTTPPKLTP 583

Db 540 RAAPPSNEATPAVPSPSPNLKSPPLPIPKVPLMALTPQPVTAQWVTOLAATKPSPIVP 589  
 QY 584 TPPEKLAPTPPEKAPAPTPPELA-----PTTP-----EPTPTTP-----EPA 622  
 Db 600 KASPK-ALMTTPPPPEGLPALAAAKLLGLSSPVASAMAKTTPRLPSPVPMASPA 658  
 QY 623 PTPP-----KAAAPNTPK-----PAPTTPKEPAPTTPKEPAPTTPKEPAP 663  
 Db 659 SLGPDARVALATNNAASPAKPEAGAGNGTLAMPKANQMAPIGAAGAAQTAPMGAH 718  
 QY 664 T-----TRKTAPT-----TLKEPAPTTPKPAKEL 690  
 Db 719 THVSPMGAGATQMSPTGANTHMSPIGAGATQMSPMGANQMSMGATQMSMG 778  
 QY 691 APPTTKP-----TSTSDKAPAPTPKG--TAPTTPKEPAPTTPKEPAPTTPKGTAP 740  
 Db 779 AATTQSPMGAAATQVTAISAGNTMOVSPGAAATTPQTPSVGAATTP-QSPM---GAA 833  
 QY 741 TTLEKAP-APTTPKPAKELAPTTKGP-----TSTSDKAP-APTTPKEP--APTTPKE 790  
 Db 834 TTLMSPGATTPQ---PSPMGAVTTQPPPMATNTTQPPPMASPTQSPMGATTTQS 890  
 QY 791 P--APTTPKAP-APTTPETPP-----PTSEVSTPTTTPKEPTTHKSPDESTPEL-- 837  
 Db 891 PPMGATTTQSPMGASTPQAPPTVAGSPPTPPPIPSPTQTSFQPSKSPPPDPKAPS 950  
 QY 838 SAPPTPALENSPKPEGVPTTKTPAATKPEMTTAKDTERLRTPEPTTAAPKMT-K 896  
 Db 951 AAAGTSPAAHVANASPEV-TAVSPA--PIGVTEASADGARISPEGTATATGPKASPA 1006  
 QY 897 ETATTEKTESKITAATTOVTSTTQDTPPFITTLTKTTLAPKVYTKTKTTTETIM 956  
 Db 1007 ATADVTEAARD-VTAAATVPA---EAAP-----TKAKRSSSSSSSSSSSSSS 1051  
 QY 957 KPEETAKPKDRATNSKATTPPKPPTKAPKPKTSTKPKTPPVKPKTTPPTPKMTSTM 1016  
 Db 1052 SSSSSSSSDSSSSSSSESNPASPA-----PVGDCQOQMTGGAOSVP 1096  
 QY 1017 PELNPTSRIAEAMLOTTTRPNOTPNSKLVENPKSEDAGAGE 1060  
 Db 1097 P-----VTEAAVQ-----EAAAAAAGAEERB 1119

RESULT 9  
 T25697  
 hypothetical protein F16F9.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T25697  
 R:Fulton, B.  
 submitted to the EMBL Data Library, August 1996  
 A:Description: The sequence of C. elegans cosmid F16F9.  
 A:Reference number: Z20071  
 A:Accession: T25697  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1229 <FULT>  
 A:Cross-references: EMBL:U67956; P1DN:AA07691.1; GSPDB:GN00028; CESP:F16F9.2  
 A:Experimental source: strain Bristol N2; clone F16F9  
 C:Genetics:  
 A:Gene: CESP:F16F9.2  
 A:Map position: X  
 A:Introns: 35/1; 364/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 9.2%; Score 674; DB 2; Length 1229;  
 Best Local Similarity 27.8%; Pred. No. 2.4e-21;  
 Matches 328; Conservative 89; Mismatches 384; Indels 380; Gaps 56;

QY 73 RCFESFGRGECDCDAQCKYDKCCPDYSEFCAEVHNPTSP-----PSKKAAPPS 123  
 Db 34 RDKIKYKHKKNTCTGSC-----KCVDPASNPFDVSTTSSINNNDVIGDSDSNPTGS 89

```

QY 124 GASOTIKSTRSRPSPKPKKTKKVKISEETIEVKONKNNRKKPKPPVVDAGSGL 183
DB 90 SMFOEIAIVGGO-----TVKSEHNIDSSVEY-----EKKVT----- 121
QY 184 DNGDFKVTPTDSTTQ---HNKYSTPKITTA-KPINRPSLPNSDT-----SKETS 232
DB 122 -----TSDASPTNAPTGGKSTPTPEITIGIVINSKSSSVYDMSTRSTLSPTTE 174
QY 233 LYNAKKEVTEKETTNNKOTSDGAKKITSAKTOSTIKTSANDLAPTSKVLAKPPKA 292
DB 175 ILTSPLELVSTDSST-EOTSPPNTIEIASPMETNTTATTTTSVPSVSTLAS--ED 230
QY 293 ETTTKGALTPKPEPTTKEPASTPKPEPTTIKSAPTTKEPAPTTPKSAPTTKE 352
DB 231 ETTVTAIAEST-----TVIAEKVSTTEPTT-----AESTTKKSTT----- 268
QY 353 PAPTTKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTKEPAPTTP---KKPAPTTPKEP 408
DB 269 KAPATTEEPPTTTEE---VTTTEASTTTSSETSTEK---PTTPLIDNKIAGPATGK-- 321
QY 409 APTTPKEP-PTTPPKPEAPTTKEPAPTTPKEPAPTTPKAPATTPKPEPAPTTP 467
DB 322 -PETTHPVLGTTPN--FDATETPEVAKSEDKMTLSKATATETTOOTTEVT-DGPEKET 377
QY 468 TKE-----PSPPTPKPEAPTTTTSAP-----TTTKEPAPT-----TTKSAPTTPKEPSP 511
DB 378 TKNVSIETPLTTPVLTSTSTFASKESDGFHTTLKLTATNDSDESATTVKPPNEE 437
QY 512 TTTKE---PAPT-----TPK-----EPAPTTPKAPAP----- 535
DB 438 TTKRSHVVPKTKGVKTVKLELSPDEPTEIT-KAPHGKLEKKTYHFLVSLDNFARY 496
QY 536 TTPKEPAP-----PAPTTPKEPAP-----TTTKKPAPPA--PKEPAPT---PK 571
DB 497 SEAKENDVYHMDYNNHREKKEPTTTEESSTTEVTTTEPANTGNPTTENPTTTOPT 556
QY 572 ETAPTPPKKLPTTPPEKLA---PTTPKEPAPTTPPEELAPTTPEPTTPTEEPAPTTPKA 628
DB 557 STASTTALPFTTEQVVTTEEPPTAKSTATO---KPTTQOSVST---EKSTTKKA 609
QY 629 AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPAK 688
DB 610 ---STTEE--PTTDEPTT---ESSTTKATTPBELSTTSEETTELKITE----- 655
QY 689 ELAPTTPKEPTTSDKAP-----TTPKGTAPTPKEPAPTTPKEPAPTTPKGTAPTTLK 744
DB 656 ---GSTTTEEPPTTAIFAASGTITIDEETSTTSTTEPITSTKE--IVTESATITOTSVS 711
QY 745 EPAPTTP-----KRAPKELAPTTTGP----- 767
DB 712 VVESPTRQPLPERKKAIVNKFKNLEVLKCKRLLKKEKESSTGSDSSSETTVVAENID 771
QY 768 --ISTSDKAPPTTPKETAPTTPKEPAPTTPKAPAPTTP--TPPTTSE-VSPPTTKE 822
DB 772 EVTTTEKAKVQPTTPITEKSTIOEETTTTTEKTSKTTTEKPTTSATTEETTSE 831
QY 823 PTTIHKSPDSTPELSEAPPKALENSPKRPGVPTTKTPATKREMTTTLKDKTENDLR 882
DB 832 PST-----TEST-----TVDTSSATTEESSTAETTTTSAE-- 862
QY 883 TPETTTA-----APKMTKEPTATTE 903
DB 863 -TSTTTTSESAAFIITGSPENTALQSSOKSEENESSAEPKARGARDVVRKKNHTTVAPAE 921
QY 904 KTESKTTATTTQ-VIST---TTODTTPFKITTLKTTTLAPKVT---TKKTIITTEIMN 956
DB 922 TTSAAVASTTTEETTEETSTLEETPIEATILNEVTGAFVYTGAVDETTITWLELIS 981
QY 957 KPBEFAPKPRATNSKATTKPO-----KPKAPKPKPTSTKKPKIMPR 999
DB 982 K-----INNTQISQKPRPIDISKTDALSSISGLIGSFKAPMAPPT----- 1022
QY 1000 VRKPKTTPTRKMTSTMPLEIN-----PTSRIAEA 1028

```

```

DB 1023 -----HTTDAFAVTAATASLNDGSDKKIIDEAOPTDELIRRA 1059

RESULT 10
A35175
mucin 1 precursor, repetitive splice form A [validated] - human
N;Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episia
ncretatic mucin; polymorphic epithelial mucin (PEM)
N;Contents: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence, revision 20-Apr-2000 #text, change 02-Jun-2000
C;Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218;
R;Jungblut, M.J.L.; Vos, H.L.; Gemlissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene
A;Reference number: A35175; MUID:9020794
A;Accession: A35175
A;Molecule type: mRNA
A;Residues: 1-952,1033-1344 <LIG1>
A;Cross-references: GB:M32738; GB:J05288; NID:9182121; PIDN:AAA35804.1; PID:9182124;
A;Experimental source: splice form A
A;Note: GenBank entries HOMEPIST1 and HOMEPIST2 present only the amino- and carboxyl
A;Accession: B35175
A;Molecule type: mRNA
A;Residues: 1-19,29-952,1033-1344 <LIG2>
A;Cross-references: GB:M32739; GB:J05288; NID:9182126; PIDN:AAA35806.1; PID:9182129;
A;Experimental source: splice form B
A;Note: GenBank entries HOMEPIST1 and HOMEPIST2 present only the amino- and carboxyl
R;Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Dunig, T.; Peat, N.; Burc
J. Biol. Chem. 265, 15286-15293, 1990
A;Title: Molecular cloning and expression of human tumor-associated polymorphic epilth
A;Reference number: A35886; MUID:90368715
A;Accession: A35886
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-992,1033-1344 <GEN>
A;Cross-references: GB:J05582; NID:9188869; PIDN:AAA59876.1; PID:9188870
A;Note: GenBank entry HUMPNMU contains one copy of the tandemly repeated sequence
R;Jan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A;Reference number: A35887; MUID:90368716
A;Accession: A35887
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A;Cross-references: GB:J05582; NID:9189598; PIDN:AAA60019.1; PID:9189599
A;Note: GenBank entry HUMPNMU contains four fewer copies of the tandemly repeated se
R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky,
Eur. J. Biochem. 189, 463-473, 1990
A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen
A;Reference number: S10571; MUID:90276413
A;Accession: S10572
A;Molecule type: mRNA
A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A;Cross-references: EMBL:X52229; NID:937053
R;Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A;Reference number: S40293
A;Accession: S40293
A;Molecule type: mRNA
A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A;Cross-references: EMBL:X52229; NID:937053; PIDN:CAA36478.1; PID:937054
R;Abe, M.; Sidiqul, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A;Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associat
A;Reference number: A36735; MUID:90088473
A;Accession: A36735
A;Molecule type: mRNA
A;Residues: 1-142,'O',144-162,'O',164-168 <ABE>
A;Cross-references: EMBL:M1823; NID:9181542; PIDN:AAA35757.1; PID:9181543
R;Masuzawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu

```





Db 2399 -SEDDLTGSGSTENIPASSKQITISSTPTDITLASEPTKSTMSDPLSTTNVLESS 2457  
Qy 624 TTPAAAPNPKKEPAPTT-----PKKEPAPTT-----TPKEAP-----TT 657  
Db 2458 TTPSSS-----KSVSSSTEGISIVSTERSKYPESTISSVLEBDLTKTTPILLEBT 2513  
Qy 658 PKETAPTPKGAFTTTPKEPAPTTPKKPAKELAPTTTKEPTSTSDKA-----P 708  
Db 2514 ASEISELTDLSLVSRINHELTSSENVKRESSTTSSSSKPSQEPAGILSTVVP 2573  
Qy 709 T-----TP--KGTAP--TTPK-----EPAPTTKEPAPTTPKGTAP 740  
Db 2574 TSSVSLITASEIEAITGNTPEFGKGRPTITSPKSLVKTSTSPNVTSSSESESKRTTTS 2633  
Qy 741 TTLKEPAPTPPKKPAKELAPTTTKGPT--STSDKPAFTTPKKEAPTTKEPAPTTKKP 799  
Db 2634 TTVSTTPTTEETTESLILTAAPSKPTSTESSEAPTPPAKTSKPSNVSTSKST 2693  
Qy 800 APPTPEPPTTSEVSTPTTTPKPTTTHKSPDESTP--EISAEPPTKALE--NSPKPEGV 855  
Db 2694 ENVTSTSGSGSLESSMSTST-----SEPETNAPAVTVSSSEASSTTLENSSTSP-- 2745  
Qy 856 PTTKTPATK-----PKMTTAKDKTTERDLRTTPETTAAPKMTKEPATTTKTESKI 910  
Db 2746 --TSSEASVLSLFPESITSEAVTVSSR---APAEITWSSESHREISTVSESEPEI 2799  
Qy 911 TATTT-----QVSTTQDTP--FKIT-----LKTITLAPKVTITTKTI 949  
Db 2800 PLSTTVSPNVVITASSISEPILSSVTSSTPRVRLITGPDDLVTVVPSHGNNRQNT 2859  
Qy 950 TTEIMNKPPEETAKPKDRATNSKATTPKPKPAKPKPTSKKPTMPRVKPK--TT 1006  
Db 2860 TASSV---PNSNSPILPSES--LTTPQPPPTTTTAKPATTSKRPPSIQPAEMFTT 2915  
Qy 1007 PTPKMTSTMPELN-----PTSRIAEMLOTTT 1034  
Db 2916 PAP-----PPSNGXGGEETNOEEQVTSITT 2942

RESULT 13  
T45025  
muscn MUC5B, tracheobronchial [imported] - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T45025  
R:Dessey, J. L.; Guyonnet-Duperal, V.; Porchet, N.; Aubert, J. P.; Laine, A.  
J. Biol. Chem. 272, 3168-3178, 1997  
A:Title: Human muscn gene MUC5B, the 10.7 kb large central exon encodes various alternat  
A:Reference number: Z22899; MUID:97166151  
A:Accession: T45025  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3570 <Des>  
A:Cross-references: EMBL:Z72496; MID:g1834502; PIDN:CAA96577.1; PID:g1834503  
A:Experimental source: Placenta  
A:Genetics:  
A:Gene: MUC5B

Query Match 9.1%; Score 663; DB 2; Length 3570;  
Best Local Similarity 26.1%; Pred. No. 1.9e-20;  
Matches 322; Conservative 92; Mismatches 489; Indels 330; Gaps 54;

Qy 54 MECCPDFKRYCTAELSCGRGFESFERGCDCAQCKKYDKCCPYE--SFCAEV-HNP 110  
Db 2279 VESLDEGLVC-----RNRE-----QVGR-KKCFNKEIVFCNIGHCP 2317  
Qy 111 TSPSSKAPPGASQT---IKSTTKRSPKPKKKTKVISEEITEVKNKKNTKK 167  
Db 2318 STPATSTAPSSSTPGTTWILTKLT-----TATTESTGTAPSS----- 2359  
Qy 168 KPPPKPPVVDAGSGLDNGFXYTTPDTSTTQINKVSTSKITTKA-----PINRP 219

Db 2360 --TGGP-----AG-----TPHVSTT-----ATPVTSSKATPPSSPGTATAP 2397  
Qy 220 SLPPNSDTSKETSGLVKNKETTVE-----KETTNNKQSTDGKEKTSKAKQSIKTS 274  
Db 2398 ALRSTATTATPSATPAIPSSIGTWTWRLSQTTPMATMSAPSSPERVHTSTVLTT 2457  
Qy 275 AKDLAPSKYLAR--TPKMETTKGALTT-----PKERTPTPKRPASTTKERTPT 327  
Db 2458 ATTTGATGSV-APSSSTPGTAHTTKVPTTTTGTGTVVSSSPGATRPVWISTTTPTT 2516  
Qy 328 IKSAPTPKKEPAPTTTKSAPTTKEPAPTT--TKKEPAPTTKEPAPTTKEPAPTTKSA 386  
Db 2517 SGSVTVTSSIPGTHHTTTLTTTQPVATGSMATPSSSTQTSPPSLITATITATGS 2576  
Qy 387 TT-PKEPAPTPKKP-APTPKEPAPTT-----TPKEPTPTPKKEPAPTTKKEPA 440  
Db 2577 TTNPSSTPGTTPPIPELTTLTATPAATSSVTPSSALGTHHTPPVMT--TATTHGRSL 2634  
Qy 441 PTAAP-----KKPAPTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPA 480  
Db 2635 PSSPHVYRTAMTSATSGTLGTHTEPSTGISHTPAAT---GTTTSTPALSPP-HPS 2689  
Qy 481 PTTTKSAPTT--TKKEPAPTTKSAFTTPKEPSPTTKEPAPTTKEPAPTTPKKAPPTPK 539  
Db 2690 SRTTESPPSGTTPPGHTTATSRRTATATPSKRTSLILSQPTSAPIIT---VVTTCG 2745  
Qy 540 EPAPTTK-----EPAPTT-----TPKKP----- 557  
Db 2746 EPQCAWSEMDLSYPMGPGSGDFDTYSNIRAGAVACEOPLGECHRAATQAPVLDELG 2805  
Qy 558 -----APPAKKEPAPTTKKEPAPTT- 577  
Db 2806 QVVECSIDFGLVGNRQVGFKFCFNYELRVCCNAGHCPSPATSPATSSTPGCTW 2865  
Qy 578 -PKKLPTPEKLAFTTPEKAPAPTTPELAPTTPEEPPTTP--EEAPTPKKAAPATPK 635  
Db 2866 ILTELT-----ATTASGSTATPSSPTGTAAPPKVLVSPATTPATSSKATSSSPR 2920  
Qy 636 EPAPTTKEPAPTTPKKEPAPTTKETAFTTPKGAFTTLEKAPTTLEKAPTTKKAPKELAPTT 695  
Db 2921 -TATTLVLNSTAKSTATSVTP-----IPSSLTGTGLPEQTTVPVATMSTIHSSST 2973  
Qy 696 KEPTSTSDKPAFTTPKGAFTTPKKEPAPTT-----TPKKEPAPTTPKGTATTLKEPAPTT 750  
Db 2974 PETHTSTVLTKATPATSTGTSSTPGCTWILTLTAATTTAGTGTATPSSPTGT 3033  
Qy 751 PKKPAKELAPTTT--KGPTSTSDKPAFT--TPKETAPTTKEPAPTTPKKPAFTTPE 805  
Db 3034 TWILTELTATTTASTGSTATLSSPTGTWILTEPSTATVAPPGSTATASSTQATAG 3093  
Qy 806 TPPTTSEVSTPTTTPKPTTIHKSPEST--PEL-SAEPTPKALENPKPEGVPT----- 857  
Db 3094 TPHVST-ATTPVTSSKATPSSSPGTAALPALRSATTPATISFT---AIPSSSLGT 3148  
Qy 858 -----TKTPAATKPEMTTAKDKT--TERDLRTTPEIT-----TAAPKTKETATTT 903  
Db 3149 TWTRLSQTTTPATMTATPSSSTPEVHTSTVLTTATTTGATGASVATPSSPGTATHTK 3208  
Qy 904 --KTESKITAAT-----TVOVSTTQDTP-----PKIT 931  
Db 3209 VPTTTTGTATPSSSPGATLTPVWISTTTTPTTPTTSGTVVPSSTIGTHTARVL 3268  
Qy 932 TLKTTTLA-----PKVTTKTITITELMNPEEY--AKPDRAATNSK 972  
Db 3269 TTTTIVATGSMATPSSSTGTPSLTTATTTATGTTNBSSTPGTTPVPLTSM 3328  
Qy 973 ATPPKQPKPAKPKPTSKKPTMPRVKPKPTTPPKKMTSTMPELNPTSRIAEALQ 1032  
Db 3329 ATT-----PAATSSKATSSSPRTATTL--PVLSTATKSTATSPPIPS-----TLMTT 3377  
Qy 1033 TTRPNO--TPNSKLVEYNPKSEDAAGAGETPH 1063

Db 3378 WYVPAQTTPMSMTIHSTSTP-----ETTH 3404

# RESULT 14

151618

nucleolar phosphoprotein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 13-Sep-1996 #sequence, revision 13-Sep-1996 #text\_change 24-Sep-1999

C/Accession: 151618; S57737

R:Cell: 108, 3339-3347, 1995

J:Cell: 108, 3339-3347, 1995

A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180

A:Reference number: 151618; PMID:96019267

A/Accession: 151618

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-990 <CAI>

A/Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921

C:Gene: xNopp180

C/Superfamily: nucleolar-cytoplasm shuttle phosphoprotein

C:Keywords: phosphoprotein

Query Match

Best Local Similarity 25.8%; Score 646; DB 2; Length 990;

Matches 260; Conservative 134; Mismatches 389; Indels 226; Gaps 46;

Db 112 SPSSKRAPPSGASQTIKSTTKRPPKPKTKVIESEITEVKNKNTKPKPT-170

Db 57 SPDAKKRRPPANGI-----PKKSAKESSESSSEDEPPAKKRAQAG 102

Qy 171 PKPVVDEAGSGLDNGCFKATTPD--ISTQHNKVSTSPKTTAKPINPRSLPNDST 228

Db 103 GKKEVY-----KAVQPKAKSSSESSSEDESEETKPKPAKRAQTPK---146

Qy 229 KETSLVANKETVETKETTNNKOTSDGKETSAKETOSIEKTSKADLAFTKVLAKP 288

Db 147 -----VAAVTTPQKKAKS-----SSSESSSESEDEAKKKQPVAKV---182

Qy 289 TPRAETTKGPAITTPKEPTTPKE---PASTTPKEPTTPKSAPTTPKEPAPTTK 344

Db 183 PPKQAVVAKAGIASNNGTAAOSSSESSSESDSPAKKTAATKTPPT--KPATPAKQAKTA 239

Qy 345 SAPTPKEPAPTTPKEPAPTTPKEPAPT--TKEPAPTT-----TKSAPTPKEPAPT---395

Db 240 GKSSSKHEDSSDEQKTAASKPKPDVYSAVPPTSVSKKTLISQGTAKAESDSS 299

Qy 396 -----TPKKAPPTTPKEPAPTTP---KEPTTPKEPAPTTPKEPAPTTPAP 444

Db 300 DSDSEEQPAKKAIVAKAASAPKPLAKAETSTDESSSESSSEDEKSSVKLGVAAP 359

Qy 445 KKPAPTTPKEPAPTTP---KEPA-----TTTKEPSPTTPKEPAPT 483

Db 360 KK-APAAP--DAKSTVPAAKASAPAKKASSSDSSSNEETTTKPAAKTTPAKSATP 416

Qy 484 TKSAPTTTKEPAPTTTSAPTPKEPSPTTPKEPAPTTPKEPA-----PTTPKKAPPTTP 539

Db 417 TSKPTPKGK--ATPTSKTPAKPPTPKTSTAKKSSSSSDSSDEETTTKPAKKTTPA 473

Qy 540 EPAPT--TPKEPAPTTPKPAAPTAKEPA--PTTPKETA-----PTTPKPLTPT 585

Db 474 KSAATPTSKTP---TNSKATPTSKTPAKPPTKTSAKKSSSSSDSSSDSKKTPA- 529

Qy 586 PEKAPPTTPKEPAPTTPPEELAPTPPEPTTPPEPAPTTP--KAAAPNTKEPAPT---640

Db 530 -KRAAKTTPAKPA-----AKTTPAKPA-----AKTTPAKPAASSTPKQVPTKES 574

Qy 641 TPKEPAPTTPKEPAPTTPKEPAPTTP--KGTAFTLKEPAPTTPKKPAKELAPTTKEPT 699

Db 575 SSSSSSSSSDEKSSKPAVKTPGKAYS-----KPVVASKPAK-----KASS 621

Qy 700 STSDKAPPTTPKGTAPTTKEPAPTT--PKEPAPTTPKGTAPTTLKEPAPTTPKKPAK 757

Db 622 SSDSDSEETTTKTPKTLTSLPAVKTLPPKKAESSSDSSDSDSEK---TPAKPAK 678

Qy 758 ELAPTTTKGP-----TSTSDKAPPTTPKETAAP--TPKEPAPTTPK-PAPTTPT 806

Db 679 SAPVNTKAPAKONKASKASCSDSSSEEGKSKOPGSKPAKAPKAPKAVVANKKD 738

Qy 807 PPTTSVSTPTTTKE--PTIHKSPDESTPELSAEPKLENSPKPEPVPTTKTPAATK 865

Db 739 PSSSSSDSSGDDKQKPKAAAKDKQAKAKAPTPKAAASSSE---DSSDESVK 795

Qy 866 PEMTTAKDPTTERDLTPPETTTAAAPKMKETATTEKTESKINATTTQVSTTTQDT 925

Db 796 AKNTNTAVSKSPV---TPKAVPAAKKESSESSSEDEKQGGKNTSTKIANST----847

Qy 926 -TPFKITTKTTTLAPKYTTK--KTITTTIMNKPEETA-KPRDRATNS---KATTPKQ 979

Db 848 -----PKMAAECSSESSSEDEGANCTSGKRRRESGTGAECAYTPE--891

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTPPTP--RKWTSMPLNPTSRIAE 1027

Db 892 -----NKKLKASPNTPPVKNKELKNTPFRRVEDIEINP--RMAD 932

```

QY 512 TTKEBAPTTPKEBAPTTPKKAPATTTPKEBAPTTPKEBAPTTPKKAPATAKEBAPTTPK 571
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 188 yyykSpppsSpppyyykSpppsSpppyyykSpppsSpppyyykSpppsSpppyyykSppps 247
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 572 ETAPTTPKKLPTPEKLAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKAAAP 631
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 248 PPPpyyykSpppsSpppsSpppyyykSpppsSpppyyykSpppsSpppyyykSpppsSpp 307
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 632 NTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 691
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 308 PSpSpppyyykSpppsSpppyyykSpppsSpppyyykSpppsSpppyyykSpppsSpppyyy 366
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 692 PTTKEPTSTSDKAPATTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 751
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 367 ---KSpPpsSpppyyykSpppsSpppyyykSpppsSpppyyykSpppsSpppyyykSpp 422
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 752 KKPAPKELAPTTTGPSTSDKAPATTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 809
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 423 SPP-----PPYVYKSpPpsSpppyyykSpppsSpppyyykSpppsSpppyyykSpp 475
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: April 26, 2002, 16:32:49  
 Job time: 764 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:55 ; Search time 62.75 seconds  
(without alignments)  
795.232 Million cell updates/sec

Title: AA7  
Perfect score: 7323  
Sequence: 1 MAMKTLPIYLILLSVFIQ.....ARATRRSGQLSKWYNCP 1361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1194	16.3	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	13.0	1664	1 SLPL_CLOTM	Q06852 clostridium
3	795.5	10.9	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	8.9	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	566.5	7.7	875	1 FPI_MYTED	Q25460 mytilus edu
6	555.5	7.6	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
7	531	7.5	620	1 EXTN_TOBAC	P13983 nicotiana t
8	533	7.3	1087	1 NFR_MOUSE	P19246 mus musculu
9	530.5	7.2	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	7.2	865	1 CPN_DROME	Q02910 drosophila
11	518.5	7.1	872	1 FPI_MYTCCO	Q25434 mytilus cor
12	509.5	7.0	662	1 MUC1_XENLA	Q05049 xenopus lae
13	503.5	6.9	1970	1 RPB1_HUMAN	P24928 homo sapien
14	498.5	6.8	1970	1 RPB1_MOUSE	P08775 mus musculu
15	498	6.8	831	1 NFR_RAT	P16884 rattus norv
16	493.5	6.7	467	1 RPB1_CRIGR	P11414 cricetus
17	488.5	6.7	826	1 SSP2_PLAYO	Q02443 plasmodium
18	488.5	6.7	1020	1 NFR_HUMAN	P10436 homo sapien
19	475.5	6.5	267	1 EXTN_MAZE	P14918 zea mays (m
20	473.5	6.5	2142	1 BAT2_HUMAN	P48634 homo sapien
21	471.5	6.4	5376	1 ZAN_MOUSE	O88799 mus musculu
22	468.5	6.4	634	1 HWP1_CANAL	P46593 candida alb
23	467	6.4	817	1 VRL1_YEAST	P37370 saccharomyc
24	454	6.2	797	1 VGLX_HSVRB	P29668 equine hepr
25	454	6.2	1161	1 Y9P_YEAST	P47179 saccharomyc
26	448.5	6.1	670	1 VES0_HUMAN	Q00130 ictaluriid h
27	448.5	6.1	1229	1 N121_HUMAN	Q09223 homo sapien
28	442.5	6.0	751	1 FPI_MYGA	O27409 mytilus gal
29	440	6.0	1794	1 FPI_SCHPO	Q10172 schizosacch
30	439.5	6.0	1083	1 T2D3_HUMAN	O00268 homo sapien
31	438	6.0	2774	1 MAPA_RAT	P34926 rattus norv
32	434	5.9	3164	1 TEGU_HSV11	P10220 herpes simp
33	432.5	5.9	439	1 XP2_XENLA	P17437 xenopus lae

34	431.5	5.9	1125	1 MAP4_MOUSE	P27546 mus musculu
35	431	5.9	3256	1 K167_HUMAN	P46013 homo sapien
36	428	5.8	1185	1 DRPL_HUMAN	P54259 homo sapien
37	426	5.8	1183	1 TRRX_RAT	O94258 rattus norv
38	425	5.8	2715	1 TRRX_HUMAN	O94258 rattus norv
39	425.5	5.8	2517	1 NCR2_HUMAN	O94618 h nuclear r
40	425	5.8	3421	1 TEGU_HSVB	P28955 equine hepr
41	424	5.8	307	1 SG83_DROME	P02840 drosophila
42	424	5.8	1251	1 Y0U3_CAEL	O09550 caenorhabd1
43	420.5	5.7	1411	1 TCOF_HUMAN	O13428 homo sapien
44	419.5	5.7	2476	1 ZAN_PIG	O28983 sus scrofa
45	418.5	5.7	907	1 VGP3_EBV	P03200 Epstein-bar

## ALIGNMENTS

RESULT	ID	Accession	Standard	PRT	AA
1	MUC2_HUMAN	Q02817	Q14878		
AC	MUC2_HUMAN	Q02817	Q14878		
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).				
GN	MUC2 OR SMUC.				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_Taxid=9606;				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine;				
RA	MEDLINE=94132002; PubMed=8300571;				
RA	Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;				
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.				
RT	Identification of the amino terminus and overall sequence similarity				
RT	to prepro-von Willebrand factor.";				
RL	J. Biol. Chem. 269:2440-2446(1994).				
RN	[2]				
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.				
RC	TISSUE=Colon;				
RA	MEDLINE=93016075; PubMed=1400449;				
RA	Gum J.R., Jr., Hicks J.W., Toribara N.W., Roche E.-M., Lagace R.E.,				
RT	Kim Y.S.;				
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located				
RT	both upstream and downstream of its central repetitive region.";				
RL	J. Biol. Chem. 267:21375-21383(1992).				
RN	[3]				
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.				
RC	MEDLINE=91358717; PubMed=1885763;				
RA	Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,				
RT	Petersen G.M., Kim Y.S.;				
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays				
RL	and polymorphism.";				
RL	J. Clin. Invest. 88:1005-1013(1991).				
CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND				
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A				
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS				
CC	AGENTS AT MUCOSAL SURFACES.				
CC	- SUBUNIT: MULTIMERIC.				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,				
CC	BRONCHUS, CERVIX AND GALL BLADDER.				
CC	- PM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR				
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).				
CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND				
CC	VARIABLE AMONG DIFFERENT ALLELES.				
CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT				
CC	OF SILKWORM HEMOCYTIN.				
CC	- SIMILARITY: CONTAINS 2 WWFC DOMAINS.				
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).				

This SMS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-slb.ch/announce/> or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).

DR	EMBL; L2198;	AAB95295.1;	-
DR	EMBL; M74027;	AAA59875.1;	-
DR	EMBL; M94131;	AAA59163.1;	-
DR	EMBL; M94132;	AAA59164.1;	-
DR	MIM; 158370;	-	-
DR	InterPro; IPR000359;	Cys_knot.	
DR	InterPro; IPR000561;	EGF-like.	
DR	InterPro; IPR002400;	GF_cysknot.	
DR	InterPro; IPR001007;	VWFC.	
DR	InterPro; IPR001846;	Vwd.	
DR	Pfam; PF00007;	Cys_knot; 1.	
DR	Pfam; PF00094;	vwd; 4.	
DR	PRINTS; PR00438;	GFCYSKNOT.	
DR	SMART; SM00214;	VWC; 2.	
DR	SMART; SM00011;	VWC_def; 2.	
DR	SMART; SM00216;	VWD; 4.	
DR	PROSITE; PS00022;	EGF_1; UNKNOWN_1.	
DR	PROSITE; PS01185;	CTCK_1; 1.	
DR	PROSITE; PS01225;	CTCK_2; 1.	
DR	PROSITE; PS01208;	VWFC; 2.	
KM	Glycoprotein; Repeat:	Signal.	
FT	SIGNAL	1 20	POTENTIAL.
FT	CHAIN	21 5179	MUCIN 2.
FT	DOMAIN	1401 1747	APPROXIMATE REPEATS.
FT	REPEAT	1401 1416	1.
FT	REPEAT	1417 1432	2.
FT	REPEAT	1433 1448	3.
FT	REPEAT	1449 1464	4.
FT	REPEAT	1465 1471	5.
FT	REPEAT	1472 1478	6.
FT	REPEAT	1479 1494	7A.
FT	REPEAT	1495 1517	7B.
FT	REPEAT	1518 1533	8A.
FT	REPEAT	1534 1556	8B.
FT	REPEAT	1557 1572	9A.
FT	REPEAT	1573 1596	9B.
FT	REPEAT	1597 1612	10A.
FT	REPEAT	1613 1635	10B.
FT	REPEAT	1636 1651	11A.
FT	REPEAT	1652 1675	11B.
FT	REPEAT	1676 1683	12.
FT	REPEAT	1684 1699	13.
FT	REPEAT	1700 1715	14.
FT	REPEAT	1716 1731	15.
FT	REPEAT	1732 1747	16.
FT	DOMAIN	4815 4886	VWC 1.
FT	DOMAIN	4924 4991	VWFC 2.
FT	DOMAIN	5075 5160	CTCK.
FT	DISULFID	5075 5122	BY SIMILARITY.
FT	DISULFID	5089 5136	BY SIMILARITY.
FT	DISULFID	5098 5152	BY SIMILARITY.
FT	DISULFID	5102 5154	BY SIMILARITY.
FT	DISULFID	? 5159	BY SIMILARITY.
FT	CARBOHYD	163 163	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423 423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	670 670	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	770 770	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	894 894	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1139 1139	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1134 1134	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1215 1215	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1230 1230	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1246 1246	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1787 1787	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1820 1820	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	4339	4339	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4351	4351	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4362	4362	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4373	4373	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4422	4422	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4438	4438	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4502	4502	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4616	4616	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4627	4627	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4752	4752	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4787	4787	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4881	4881	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4888	4888	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4955	4955	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	4970	4970	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	5019	5019	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	5038	5038	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	5069	5069	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CONFLICT	1351	1351	H -> L (IN REF. 3).		
FT	CONFLICT	1412	1412	T -> S (IN REF. 3).		
FT	CONFLICT	1449	1449	L -> P (IN REF. 3).		
FT	CONFLICT	1504	1504	M -> T (IN REF. 3).		
FT	CONFLICT	4192	4192	G -> S (IN REF. 2).		
SQ	SEQUENCE	5179 AA:	540295 MM:	85CD7571FB9A5663	CRC64:	

Query Match	16.3%	Score 1194:	DB 1:	Length 5179:
Best Local Similarity	30.6%	Pred. No. 5.6e-40:		
Matches 400:	Conservative	74:	Mismatches 515:	Indels 320: Gaps 51
QY 4	KTLPIYILLLLSVFIQVSSODLSSC--AGRCG-----EGYSRDATNCNDYNCQHVMEC	56		
DB 1169	KRPRIY-----EDLKKCVIADKCGCYVEDTHYPGASVPTFEKSCV-C	1213		
QY 57	CPDFKRVCTAELSCGRCSEFPERG-----RECDCAOCKKTKDKC-----PDYESFCAE	106		
DB 1214	TNSSGVYCRPE--EGKILNQTODGAFCCYMEILCGPVGTEVKEHFNICISITRSTLTTFPT	1270		
QY 107	VHNTPSPS-SKKAPPSGASQITKSTKRSPPPKKKTKVKKIESEETFEVANKNRT	165		
DB 1271	ITLPIPTFTTTTTTTTTPPSSIVLST--PK-----LCLMSDWLNE-----	1311		
QY 166	KKKPKPKPVYDEAGSGLDNGDFK-----VTPDPTSTQH-NKVSITRK	208		
DB 1312	-----DHPSSGSDGDGREPDVGAGCAEDIEDCKSVKRNHSLDGHGKVCDDVS	1360		
QY 209	-----TTAKINIRPSLSPNSQTSKETSL	233		
DB 1361	VGFICKNEDQFGNGPFLCYDKIRVNCNCMPMDKCIITPSPPPTTSPPTTTTLPLPT	1420		
QY 234	TVNKEETVEIKENTNTNKOTSTDGKEKKTSAKETQSIEKTSADKLAPYSKVLAKPTPKAE	293		
DB 1421	TPSPPT-----TTTTPPPTTTPSPPTTTTTTP-----LPTT-----TPSPIS	1459		
QY 294	TTTSGPALTPKEPT-----PTTPKEBASTTPKEAPTITKAP--TTPKEP-----AP	340		
DB 1460	TTTTPPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTPPTTPSPSPMTPIIPPASTTLPP	1519		
QY 341	TTTTKAP--TTPKEBAPTTPKEAPTTPKEP-----APTTPKEBAPTITKAP--TTPK	390		
DB 1520	TTTTSPPTTTTTTPPTTTPSPPTTTPPTTTPPTSTTLPLPPTTSPSPPTTTTTTTPPTTSS	1579		
QY 391	EPAPPTPKKAPPTPKKAPATTPKEAPTTPKEAPTTPKEBAPTTPKEAPTTPKEP-----APTA	443		
DB 1580	PPPTTTPSPPTTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTLPTT	1639		
QY 444	PKKPAAPTTPKEAPTTPKEBAPTTPKEBSPTPKEBAPTITKSAPTTTPKEBAPTITKAP	503		
DB 1640	TPSPPTTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPSPPTTTPSPPTTMTPT	1699		
QY 504	-----TTTKEBSPITTKAPPTTPKEBAPTTPKKKAPPTTPKEBAPTTPKTKKAP	560		
DB 1700	SPPTTTPSSPI-TTTTTTPSSITTPSPPTTMTTPSPPTTTPSPPTTTPPTTMTTLPLPTTSSPLTT	1758		

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL; X67506; CAA47841.1; -
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH DOMAIN; 2.
KW Cell wall; S-layer; Signal; Repeat.
FT SIGNAL 1 28
FT CHAIN 29 1664
FT DOMAIN 36 763
FT REPEAT 36 191
FT REPEAT 207 363
FT REPEAT 409 565
FT REPEAT 607 763
FT DOMAIN 771 1377
FT FT 1378 1449
FT FT 1453 1494
FT FT 1495 1565
FT FT 1566 1625
FT FT 1626 1646
FT FT 1664 AA; 178194 MW; 5F39695BA9FE74B CRC64;
SQ SEQUENCE

Query Match 13.0%; Score 950; DB 1; Length 1664;
Best Local Similarity 31.0%; Pred. No. 7.9e-31;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

OY 284 VLAKPTP-KAETTTKGALTTPKKEPTPTTPKKEPASTTPKEPTTIKSAPTTPKKEPAETT 342
D 758 VVIGAPAKKAASDEEPIPTDPSDEPTPS-----DEPTPS--DEPTPSDEPTPSD 804
OY 343 TKSATTPKEPAPTTPKKEPAPTTPKEPA-----PTTKKEPAPTTTSATTPKKEPAPTTPK 398
D 805 EPTPSETPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 864
OY 399 KPAPTTPKEPAATTPDKEEPTPTTPKKEPAETTPKEPAPT--TPKKEPAAPKAPATTPKKEPA 457
D 865 SDEPTPSDEPTPS--DEPTPS--DEPTP-SDEPTPSETPEEPIPTDPSDEPTPSDEPTP 919
OY 458 TTPKKEPA--PTTTPKPESTP--TPKKEPAPTTYS-APTTPKEPAPT--TTKSAPTTPKKEPS 510
D 920 SDEPTPSDEPTPSDEPTPSSETPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 979
OY 511 ----PTTKKEPAPT--TPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAAPKAP 565
D 980 PSDEPTPSDEPTPSSTPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1039
OY 566 APT--TEKETAPTTPKKAATTPTPPEKLAATTPPEKAPAT--TPBELATTEDEPTPT--TP 618
D 1040 TPSDEPTPSE-----TPPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSD-ETPSDEPT 1094
OY 619 EEPATTPKKAAPNTPKKEPAATTPKEPA--PTTPKEPAPT--TPKTAATTPKGAATTPLK 675
D 1095 EEPITDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1154
OY 676 EEPATTPKKAAPKELAPATTPKKEPT--STSDKAPATTPKGAATTPPEKKEPAATTPKKEPAT 734
D 1155 EPTPS--DEPTPSD-EPTPSDEPTPSSETPEEPIPTDPSDEPTPSDEPTPS--DEPTPS- 1208
OY 735 PKGATPPTLKKEPAPT--TPKKAPKELAPATTPKGTSTTSKAPAP--TPKETAPTTPKE 790
D 1209 ---DEPTPSDEPTPSSETPEEPI-----PDTPSDEPTPSDEPTPSD-ETPSDE 1253
OY 791 PAPTPPKKAPAT--TPENPTPTTSKVSPTTTKKEPTTIKSDSETPPELSAETPKALENS 849
D 1254 PAPTPPKKAPAT--TPENPTPTTSKVSPTTTKKEPTTIKSDSETPPELSAETPKALENS 849

```

OY	850	KPEBGVPTTKT	PAAT--KPMITTAAKD	TENDLRTPETTTAABRKMKETAATTTEKTE	907
Dd	1306	SDEPSPDEP	PSETPSEEPPIPDTPSDEPPSD---	EPPSPDEPSPSEPTSPDEPPPS	1361n
OY	908	SKIAATTOYTASTTOOTPFKTTTLTKTTTAAPKVTTTKTTTTTEIMNKKPETAKPKDR			967
Dd	1362	DEPPSPDEPPTSPSETPEEPT-----	TTTTPTTPTST-----	-----PTSG	1366n
OY	968	ATNSKAT-----	TFKPQKPTPKAB-KKPSTKKKKTPRVRKKKTPTPRKM-----		1012n
Dd	1397	SGSGGGGGGGGGGCGVPTSPPTPTSKTSPAP---	TEIEBTPSDVPALIGEHA	1453n	
OY	1013	-----	TSIMEPLNPSTRAEAMLOTTTRPNOTPMSKI.VEVNPKSESADAGACGETPHM-	1064n	
Dd	1454	YLRCGYPDGSFPERENITRAEAAVIF-----	AKLL-----GADESQASASPYS	1498n	
OY	1065	LLRPH-----	VEMFE-----VFEDMOYLPVRPOGI-	1090n	
Dd	1499	LADTHMAAMAIIKFAITSOGLEKGYPDGTFFKDONITRAEFATIVYLHLTKVKGOELMSKLA	1556n		
OY	1091	---IINMLSDNETNICNG---KPVDGILTYL-----	RNGTLVAERGHYFWMLSPFSBPS	1137n	
Dd	1559	TIDISNRPFD-----CVGHMAAQEIFIELKINSIGVISGYPGDT-----	FKQN	1600n	
OY	1138	PARITEWYGIPSPIDIYVTRFCMECEGTKTFEKO--SOYRRFTNDIKD	1182n		
Dd	1601	YIKKSESY---ALINRALRGRLNAPKLFPDVNESVAF-GDIMD	1642n		

RESULT 3  
AMTH\_YEAST STANDARD; PRT: 1367 AA.  
ID AMTH\_YEAST P08640; P08640; (Rel. 08, Created)  
AC 01-AUG-1988 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 40, Last annotation update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GLUCOSYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-  
GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).  
GN STR1 OR STR2 OR MAL5 OR YIR019C.  
GN Saccharomyces cerevisiae (Baker's yeast).  
OS Saccharomycetes: Ascomycota: Saccharomycotina: Saccharomycetes;  
OC Eukaryota: Fungi: Ascomycota: Saccharomycetales: Saccharomycetes.  
OC Saccharomycetales: Saccharomycetaceae: Saccharomycetes.  
OX NCBI\_Taxid=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
RA Churchill C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
RA Gents S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
RA Walsh S.V., Whitehead S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.  
RC MEDLINE=87194600; PubMed=3106330;  
RA Yamashita I., Nakamura M., Fukui S.;  
RT "Gene fusion is a possible mechanism underlying the evolution of  
RT STR1.";  
RL J. Bacteriol. 169:2142-2149(1987).  
RN [3]  
RP SEQUENCE OF 1-31 FROM N.A.  
RC STRAIN=SPX101-1C;  
RX MEDLINE=69031230; PubMed=3141213;  
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;  
RT "Similar short elements in the 5' regions of the STR2 and SGA genes  
RT from Saccharomyces cerevisiae.";  
RL FEBS Lett. 239:179-184(1988).  
CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-  
CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS  
CC WITH RELEASE OF BETA-D-GLUCOSE.  
CC -I- SIMILARITY: TO S.POMBE SPBC315.13.  
CC

```

CC      -I- SIMILARITY: SOME. TO S.POMBE SPCC285.13C.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z38061; CAA86176.1;  -
DR      EMBL; M16164; AAA35014.1;  -
DR      EMBL; M16165; AAA35015.1;  -
DR      EMBL; X13857; CAA32069.1;  -
DR      PIR; B26877; B26877.
DR      PIR; A26877; A26877.
DR      PIR; S48478; S48478.
DR      SGD; S0001458; M0C1.
KW      Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW      Signal; Multigene family.
FT      SIGNAL. 1 21
FT      CHAIN 22 1367
FT      DOMAIN 210 1367
FT      CAROXYD 817 817
FT      CAROXYD 874 874
SQ      SEQUENCE 1367 AA; 136110 MW; 91C0E3DBD61AA9D CRC64.

```

Query Match	10.9%	Score 795.5	DB.1:	Length 1367
Best Local Similarity	27.4%	Pred. No. 7.6e-25		
Matches 318:	Conservative 111:	Mismatches 525:	Indels 207:	Gaps 51:
OY	85	DDDAQCKRKDK----	CCPDYSEFCAEVHNHPTSPSSKKAPPPGASQJTKSTTKRSPKPPN	141
DB	174	DISTGCNNYDNGHSGHOTDEPGFYNNHDDCNDNCGGKRSSTTTSSTSES--	STTTSS-----	226
OY	142	KKRTKVKIESEIEIVKDNKKNRKKKPKPKPVYDEAGSGLDNGDFKVTTPDITQHN	201	
DB	227	-----TSEESTTSSSTSESST-----	TSTSTSESTSS	253
OY	202	KVSTSPKI--TTKKPINPRLSPNBDTSKETSILYANKETTVKETTNTKQTSIDGKEK	260	
DB	254	SSSTIAPAPPTTISCTREKERTPTTTSCTREKERTPPHHDTPTCKTKTTTSK--TCI--	KK 309	
OY	261	TTSANETOSIEKTSADLAPTSKVLAKPPTPKAETT--TKGALITTPKE-----	PTPTTP 312	
DB	310	TTTTPTPTPS--SSTTESSSAPV-----	PTSSSTTESSSAPVTSSTTESSSAPVPPTSSS	362
OY	313	KRPASTTREPRTPTIKSAP----	TTKREPATTTKSAPTTPKEAPTTTKAPATTKE 368	
DB	363	TTSESSAPVTSSTTESSSAPVTSSTTESSSAPVPPTSSSTTESSSAPVTS--	TTSSS 418	
OY	369	PAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKAPATTTPKEPATTKEPPT--	TTPKKPA 425	
DB	419	SAPVTS--STTESSSAPVT-----	STTESSSAPVTSSTTESSSAPVPPTSSSTTESSSA 471	
OY	426	PTTKKEPAPTTTKEPAPTPADKKAPPTTPKEPAP----	TTPKKEPAPTTTKEPSPPTTTPKEPAP 481	
DB	472	PVT-----SSTTESSSAP--VPPTSSSTTESSSAPVTSSTTESSSAPVPTPS--	SSTTTSSSAP 526	
OY	482	TTTKSAPTTPTKKEPAP-----	TTTKSAPTPPKPEPPTTTPKEPAP-----	TTPKKEPAP--TPKK 532
DB	527	APTPTSSSTTESSSAPVTSSTTESSSAPVPPTPSSTTSESTTPTVTSSTTESSSAPVPPTSS	586	
OY	533	PAPTTTKEPAPTTTKEPAPTTTKKAPTAAPKAPADPTTKEPATTPTPKKJPTTPEKLAPT	592	
DB	587	STTESSSAPVPT-----	PSSTTESSSAPAPPTSSSTTESSSAPVT-----	SSTTESSSAP- 637
OY	593	TPKCAPPTTPEKLAPT--TPKEPTTTPKEPAP-----	TTPKAANNTKEPAPTTTPKKEPA 646	
DB	638	VPPTSSSTTESSSAPVPPTPSSTTTESSSAPVPTPSSTTTESSSAPVT-----	SSTTESSSA 693	
OY	647	PTTPKEPAPTTTKEPAPTT--TPKGIAPTLTKEPAPTTPKKPAKELAPTTTKEPPTSTTSDK	705	



RA Lee L.N., Luh K.T., Wu C.W.;  
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and  
 tissues.";  
 RN Oncology 53:118-126(1996).  
 RC  
 RP SEQUENCE OF 1-46 FROM N.A.  
 RC TISSUE-Breast carcinoma;  
 RA Bulwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;  
 Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
 CYTOSKELETON.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM  
 IS ALSO PRODUCED.  
 CC -1- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE  
 SPLICING.  
 CC -1- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL  
 TUMORS, SUCH AS BREAST CANCER.  
 CC -1- PPM: HIGHLY GLYCOSYLATED (N- AND O-LINKED CARBOHYDRATES AND SIALIC  
 ACID).  
 CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT  
 VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE  
 MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.  
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: J05582; AAA60019.1; -  
 DR EMBL: M32738; AAA35804.1; -  
 DR EMBL: M32739; AAA35806.1; -  
 DR EMBL: J05581; AAB59876.1; -  
 DR EMBL: M61170; AAB53150.1; -  
 DR EMBL: X52229; CA364478.1; ALT\_SEQ.  
 DR EMBL: X52228; CA364477.1; ALT\_SEQ.  
 DR EMBL: M35093; AAB59612.1; ALT\_SEQ.  
 DR EMBL: Z17324; CA78972.1; -  
 DR EMBL: Z17325; CA78973.1; -  
 DR EMBL: M31823; AAA35757.1; -  
 DR EMBL: S81781; AAD14376.1; ALT\_INT.  
 DR EMBL: S81736; AAD14369.1; ALT\_INT.  
 DR EMBL: M21868; AAB59874.1; ALT\_SEQ.  
 DR PIR: A35175; A35175.  
 DR PIR: B35175; B35175.  
 DR PIR: S10218; S10218.  
 DR GLYCOSULEDB: P15941; -  
 DR MIM: 158340; -  
 DR MIM: 113720; -  
 DR InterPro: IPR000082; SEA.  
 DR Pfam: PF01390; SEA; 1.  
 DR SMART: SMO0200; SEA; 1.  
 DR PROSITE: PS50024; SEA; 1.  
 KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  
 KM Repeat: Alternative splicing.  
 FT SIGNAL 1 23  
 FT CHAIN 24 1255  
 FT DOMAIN 24 1162 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1163 1186 POTENTIAL.  
 FT DOMAIN 1187 1255 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 81 960 44 X 20 AA TANDEN REPEATS.  
 FT DOMAIN 1034 1151 SEA.  
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 975 975 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1133 1133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 19 19 T -> TATTAAPKAT (IN ISOFORM B).  
 FT VARSPLIC 20 22 MISSING (IN ISOFORM C).  
 FT VARSPLIC 20 31 MISSING (IN ISOFORM D).  
 FT VARSPLIC 20 31

FT VARSPLIC 126 905 MISSING (IN POLYMORPHIC EPITHELIAL  
 FT VARSPLIC 1077 1087 ISOFORM).  
 FT VARSPLIC 1088 1255 FLUOROGEL -> VSLGSLPMLP (IN SECRETED  
 FT VARSPLIC 1088 1255 ISOFORM).  
 FT VARSPLIC 1088 1255 MISSING (IN SECRETED ISOFORM).  
 FT CONFLICT 2 2 T -> A (IN REF. 11).  
 FT CONFLICT 134 134 P -> Q (IN REF. 9).  
 FT CONFLICT 154 154 P -> Q (IN REF. 9).  
 FT CONFLICT 1021 1021 S -> T (IN REF. 3).  
 FT CONFLICT 1251 1251 A -> T (IN REF. 3).  
 SQ SEQUENCE 1255 AA; 122072 MW; 5E2BDC4DEF7D9A82 CRC64;

Query Match 8.9%; Score 651; DB 1; Length 1255;  
 Best Local Similarity 27.8%; Pred. No. 3,2e-19;  
 Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;  
 231 TSLTVNKETTVETTKETTTKNTSDGKEKTSKAKTSKTSKAKD-LAPTSKVLAKPT 289  
 16 TVLTV-----VTSGHASSTPGKEKETSATORSVSSREKNAVMTSSVLSHS 65  
 290 P-KAETTKGP--ALTTPKEP-----TPTKREPASTTPKKEPTPTT---K 329  
 66 PGSGSSTTGQDVTAPATPEPASGSAATWGQDVTVPTRPALGSTPPAHADVTSADNK 125  
 330 SAPTPKEPAPTTKSAPTPKKEPAPTTKP-----APTPKEPAPTTKKEPAPTTK 383  
 126 PAPSSTAPPAHGVTSAPDT--RAPSTAPPAHGVTSADT--RAPGSTAPPAHGV- 179  
 384 SAPTPKEPAPTTP-----KKPAPTTKPEP-----APTPKEPTPTTP----- 421  
 180 SAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAPPAHGV 239  
 422 ----KEPAPTTKP-----APTPKEPAPTAPK-----KPAPTPKEPAPTTPK-----E 462  
 240 SAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAPPAHGV 299  
 463 PAPTPKEPSTTPKEPAPTTK-----TKSAPTPTKKEPAPTTKSAPTPKEPSTTPKEP-- 517  
 300 SAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAPPAH 356  
 518 ----APTPKEPAPTTK-----KKPAPTTKPEP-----APTPKEPTPTTP----- 565  
 357 GVTAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAPPAH 416  
 566 APTPKEPAPTTKKLPPTPEKLPATTPPEKPAPTPEELAPTPPEPTTPPEPAPTT 625  
 417 GVTAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAP-----PAHGVTSAPDTRRAPGSTA 472  
 626 PKA-----AAPTPKEPAPTTPK-----EPAPTPKEPAPTTPKETAPTTPKGTAPTTKEP 677  
 473 PPAHGVTSAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAP-----PAHGVTSAPDTRRAP 528  
 678 APTP-----KKPAPELAPTTTKEPTSTSDKRAP--TPPKGTATTPKEPAPT 725  
 529 GSTAPPAHGVTSAPDTRRAPGSTAP--PAHGVTSAPDTRRAPGSTAPPAHGVTS---AP 583  
 726 TPKEPAPTTPKG-----TAPTTKEPAPTTP-----KKPAPELAPTTTKEPTST 771  
 584 TRAPAGSTAPPAHGVTSAPDTRRAPGSTAPPAHGVTSAPDTRRAPGSTAP--PAHGVTSAP 642  
 772 SDKRAP--TPK-----ETAPTPKEPAPTTP-----KKPAPTTPETPTPTTSEVS 815  
 643 DTRAPAGSTAPPAHGVTSAPDTRRAPGSTAPPAHGVTSAPDTRRAPGS--TAPPAHGVTS 700  
 816 TPTTTKEPT-----THKSPDESTPELSAPTPKALENSKPEGVPTTTPATKPEM 868  
 701 APDTRRAPAGSTAPPAHGVTSAPDTRRAPGSTAPPAHGVTSAP-----DTRAPGSTAP-- 753  
 869 TTKKDKTTEEDLTPTPTT-----TAAPKMTKETAATTEKTESKITATTOVSTT 921  
 754 --PAHGVTSAPDTRRAPGSTAPPAHGVTSAP-----DTRAPAGSTAPPAHGVTSAPDTRRA 807

QY 922 TODTP--FKITLTKTTTLPAPVTTTKKITT--ELMKNPEETAKKDRATNSKATTPK 977  
DB 808 PGSTAPPAHGVTSADTPRAPGSTAPPAHGVTSADTPRAPGSTAPPAHGVTSADTPRA 867  
QY 978 POKPTAPKPKPTSTKPKTPRVRKPKRTTPTPKMTSTMEELPTSTKIAEAMLOTTTRN 1037  
DB 868 P--GSTAPPAHGVTSADTPR--APGSTAPPAHGVTSADTPRA-----PG 909  
QY 1038 QT--PNKSLVEVNPKESEDAGSETPHMLLRPHVMEVTPMDYLRVPNOGIINPML 1096  
DB 910 STAPPAHGVTSADTPRAPGSTAPPAH-----GVTSAPDNKRALGSTA----PPV 955  
QY 1097 SDETNICNKPVDGLTTLRNGTLVAFRGHVFMLSPSPSPS 1137  
DB 956 HNVTSASGSASSTLVHNGTSARATTTTPASKSTSPSPS 996

RESULT 5  
FPI\_MYTED STANDARD; PRT; 875 AA.  
AC 025460;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT  
DE PROTEIN 1) (MEPPI) (FRAGMENT).  
GN FPI.  
OS Mytilus edulis (Blue mussel).  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
OC Mytiloidae; Mytilidae; Mytilus.  
OX NCBI\_Taxid=6550;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91025829; PubMed-1367451;  
RA Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;  
RT "Structural and functional repetition in a marine mussel adhesive  
RT protein.";  
RL Biotechnol. Prog. 6:171-177(1990).  
RN (2)  
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.  
RX MEDLINE-83135732; PubMed-6298211;  
RA Waite J.H.;  
RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and  
RT hydroxyproline-containing decapeptide in the adhesive protein of the  
RT mussel, Mytilus edulis L.";  
RL J. Biol. Chem. 258:2911-2915(1983).  
CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
CC -1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-K IS POST-TRANSLATIONALLY  
CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE  
CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE  
CC (DOPA) DERIVED FROM TYROSINE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X54422; CAA38294.1; -  
DR InterPro: IPR002964; Adhesive\_plaq.  
DR InterPro: IPR002965; P\_fich\_extensin.  
DR PRINTS: PRO1216; ADHESIVEI.  
DR PRINTS: PRO1217; PRICHEXTENSIN.  
DR Repeat: Hydroxylation.  
FT NON\_TER 1

FT DOMAIN 67 870 TANDEN REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-  
FT P-P-[ST].  
SQ SEQUENCE 875 AA; 100412 MW; 6EA85312748CACE CRC64;  
Query Match 7.7%; Score 566.5; DB 1; Length 875;  
Best Local Similarity 27.2%; Pred. No. 4.9e-16;  
Matches 274; Conservative 103; Mismatches 404; Indels 227; Gaps 54;  
QY 108 HNPSTPPSSKAPPPSGASOTIKSTKSPKPPKKTAKVISSEETEVKDNKKNTKK 167  
DB 3 HEVYKP-----VKTSYAPKPPYQPLKKKVDYRPRTKSPYVSGSKTNY 48  
QY 168 KPTPKPVVDEAGSGLDNGDFKVTTPDTSTGNKKVST-----SPKTTAKPINPRESLP 222  
DB 49 LPLAK-----KLSSYKPIITTYNAKNTNPPVYKPKWTTPPYTKKPSYP 92  
QY 223 PMSDTSKETSILTVNKETTVETKETTNNKOTSDGKEKTSIAKETOSIEKTSAKDLAPTS 282  
DB 93 P-----TKSKRTYKPKIT-----YPTTKAKPSYSSKPKKTYPTPY 131  
QY 283 KYLAKPKPAETTTKGPALTTKEPTPTT-----PKEPASTPKKEPTTTKSAPTTP-- 335  
DB 132 K-----PKLTY-----PPTKPKPSYPPTKPKPSYPPTKPKPSYPPTKPKPSYP 181  
QY 336 --KEPAPTTKSAPTPKKEPAPTTKEPA--PTPKKEPAPTTKEPAPTTKSAPTTPKEP 392  
DB 182 KAKPSYPTTKAKPSYP-----PTTKAKPKAPPTYPSTYKAKPSYPTTKAKPSYPTTKAKP 237  
QY 393 A--PTTPKKPA--PTPKKEPA--PTPKKEPT--PTPKKEPAPTTKEPAPTTKEPAPTPAK 445  
DB 238 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 293  
QY 446 KPA--PTPKKEPA--PTPKKEPA--PTTKKEPS--PTPKKEPA--PTTKKAPPTTKKEPA 495  
DB 294 KPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 353  
QY 496 PTTKSAPTPKKEPPTTKKEPAPTTKEPAPTTKPA--PTTKKEPA--PTTKKEPA--PTTKKEPAP 551  
DB 354 PSTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 409  
QY 552 TTTKKEPAPTKKEPAPTTKEPAPTTKPLTPTTPEKLABTPEKPAPTTPEELAPTPPE 611  
DB 410 PPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 457  
QY 612 EPTPTPEEPAPTTKPAANTPKKEPAPTTKEPAPTTKE-----PAPTTPKEAPTT-- 665  
DB 458 KPT-----YPTYKA-----KPSYPAKYKAKPSYPTTKKSSYSSYKPKKTYPTPY 506  
QY 666 PKGTAPTTKEPAPTTPKKPAKELAPTTTKEPTSTSDKPAPTTPGTAPTT-----PKE 721  
DB 507 PKLTKAPTYK--PKPSYPSKPKKTYPTPYK-----KRISYPTTKAKPSY 551  
QY 722 PAPTTPKEPAPTTPKGT--APTTLKEPA--PTTKKPA--PRELAPTTTKG--PTSTSD 773  
DB 552 PATTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 611  
QY 774 KPAPTTPKEPAPTTKEPA--PTTKKPA--PTT-----PEPTPTTSVSTPTTKE----- 822  
DB 612 KAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 667  
QY 823 PTTIHKSPDESSTELSAEPPKALENSPK-----EPGVPTT--KTPA-----ATKPEMTT 871  
DB 668 PPT--YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 726  
QY 872 AKCTTERDLRTPTETTTAAPKMTKETATTTKETSKITATTTQVTSSTTODTTPPKIT 931  
DB 727 YKAKPSYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 777  
QY 932 TLKTTTLAPVTTTKKTTTTLTLMNKEETAKKDRATNSKATTPKQKTAKPK--P 968  
DB 778 TYKSKSYSSYKPKKTYPTT--YKPKLTYPTPYK-----PKPSYPSYKPKKTYPT 826

```

OY 989 TSTFKKPTMRVRKKTTPP-----RKMTSMDELNPSTR 1024
DB 827 STYKLKPSPTTKSTYPTTKKTSYPSSTYKAKTSYPAKPTNR 874

RESULT 6
ZAN_HUMAN STANDARD; PRT; 2700 AA.
ID ZAN_HUMAN
AC Q9Y493; 000218;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
GN ZAN.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE OF 1-2379 FROM N.A.
RX MEDLINE=99018118; PubMed=9799793;
RA Glocker G., Scherer S., Schaltevoy R., Boright A., Weber J.,
RA Tsui L.C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes."
RL Genome Res. 8:1060-1073(1998).
RN [2]
RP SEQUENCE OF 2338-2700 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN)."
RL Genomics 41:119-122(1997).
CC - FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC - SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC - TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC - DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC - DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC - DOMAIN: THE WMPD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC - SIMILARITY: CONTAINS 4.5 WMPD DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF053356; AAC78790.1; -
DR EMBL; U83191; AAC51208.1; -
DR MIM; 602372; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000988; MAM.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR002919; TIL.
DR InterPro: IPR003328; TILA.
DR InterPro: IPR001007; WMPD.
DR InterPro: IPR001846; Wvd.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 4.
DR Pfam; PF02345; TILA; 4.

```

```

DR Pfam; PF00094; Wvd; 4.
DR PRINTS; PR01217; PRICEXTENSU.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; WVC; 1.
DR SMART; SM00011; WVC_def; 3.
DR SMART; SM00216; WVC; 3.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS50060; MAM_2; 4.
KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
FT NON_TER 1
FT DOMAIN <1 109 MAM 1.
FT DOMAIN 112 136 MAM 2 (PARTIAL).
FT DOMAIN 161 326 MAM 3.
FT DOMAIN 322 446 MAM 4.
FT DOMAIN 483 951 66 x HEPTAPEPTIDE REPEATS (APPROXIMATE)
(MUCIN-LIKE DOMAIN).
FT DOMAIN 953 1065 WMPD 1 (PARTIAL).
FT DOMAIN 1066 1454 WMPD 2.
FT DOMAIN 1455 1861 WMPD 3.
FT DOMAIN 1862 2292 WMPD 4.
FT DOMAIN 2293 2684 WMPD 5.
FT DOMAIN ? ? EGF-LIKE.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1618 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1832 1832 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1878 1878 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2136 2136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2505 2505 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 2374 2379 NOKMA -> RAQGP (IN REF. 1).
FT NON_TER 2700
SQ SEQUENCE 2700 AA; 293013 MW; 80E60C0B12277B1 CRC64;

```

```

Query Match 7.68; Score 555.5; DR 1; Length 2700;
Best Local Similarity 32.68; Pred. NO. 3.3e-15;
Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;

OY 214 PINRPSLPSPDSKETSITVNNKETTVEETKTTTQKQSTJDKERTSAKETSIEKT 273
DB 445 PVKVLPELPSPVSS-----TGSETTGLENPTLSK-----KPTVSIKRP 487
OY 274 SAKDLAPTSKVLAKPTKKAETTKGPAITTKPEPTPTTKKBPASTTKPEPT 328
DB 488 SVTTEKPT-----VPKKEKPTIPTEKPTISTEK--PTIPSEKPMPSSEKPTIPSEKPTIL 539
OY 329 KSAPTTKKBPAPTTTKSAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKKBP 387
DB 540 TEKPTITSE-KPTIPSEKPTISTEKPTVPIEE--PTTPEETTTTVMEEPAVITPKPSIPT 596
OY 388 TPKEPAPTTKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTAKRP 447
DB 597 --EKPSIPTEK--PTISMETIISTEKPTICPEKPTIPTEK--PTIPSEKTSISPEK 647
OY 448 APTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKSAPTTPK 507
DB 648 -PTTPTTE-KPTIPTEKPTISTEKPTIPTEK-PTISPEKTLIPTEKLTPT--EKPIIPT 701
OY 508 EPSPTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAP 562
DB 702 EKPTISTEE--PTTPEETTTISTEKPSIPMEKPLPLPTPEETTTISVEETTTISTEKLTIM-- 757
OY 563 KEPAATTKKETAP-----TPKKLTPTTPKLA-----PTTPEKBPAPTTPEELAPTTPEE 612
DB 758 EKPTISTEKPTIPTEKPTISTEKLTPT-EKLTIPTEKPTIPTEKLTISTEKL--TIPTE 814
OY 613 PTTTPEBPAPTTTKAAPTNTKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKGAP 672
DB 815 KPTIPSEKPTISTEK-----PTIPTE-KPTIPTE-----ETTISTEKLTIPT 855

```







RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,  
 RA Prioli R.P.;  
 RT "The Trypanosoma cruzi neuraminidase contains sequences similar to  
 RT bacterial neuraminidases, YWTD repeats of the low density lipoprotein  
 RT receptor, and type III modules of fibronectin."  
 RL J. Exp. Med. 174:179-191(1991).  
 RN [2]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE-91376547; PubMed-1896773;  
 RA Prioli R.P., Mejia J.S., Alj T., Akawa M., Pereira M.E.A.;  
 RT "Trypanosoma cruzi: localization of neuraminidase on the surface of  
 RT trypomastigotes."  
 RL Trop. Med. Parasitol. 42:146-150(1991).  
 CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN  
 CC PARASITE INVASION OF CELLS.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC  
 CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACETYLNEURAMINYL  
 CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETLATED  
 CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,  
 CC GLYCOLIPIDS OR COLOMINIC ACID.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
 CC (POSSIBLE).  
 CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM  
 CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.  
 CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT  
 CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA  
 CC PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.  
 CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR EMBL: M61732; AAA30255.1; -  
 CC DR PIR: JH0557; JH0557.  
 CC DR HSP: P29768; IDIU.  
 CC DR InterPro: IPR002860; BNR.  
 CC DR Pfam: PF02012; BNR; 2.  
 CC KM Hydrolyase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;  
 CC KW Phosphorylation.  
 CC FT DOMAIN 1 457 CYS-RICH.  
 CC FT REPEAT 23 34 BNR 1.  
 CC FT REPEAT 163 174 BNR 2.  
 CC FT REPEAT 209 220 BNR 3.  
 CC FT DOMAIN 458 588 FIBRONECTIN TYPE-III.  
 CC FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS. LTR DOMAIN.  
 CC FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 7.2%; Score 530.5; DB 1; Length 1162;  
 Best Local Similarity 29.4%; Pred. No. 1.6e-14;  
 Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;

OY 308 TPTPKPE-PASTPKPEPTTTKSAPTTTPKEPAATTTKSAPTTTPKEPAATTTTPK 366  
 DB 599 TPTPKPE-PASTPKPEPTTTKSAPTTTPKEPAATTTKSAPTTTPKEPAATTTTPK 366  
 OY 367 KE-PAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 413  
 DB 652 ADSSAHGTSTPTVDSAHSTPTVDSAHSTPTVDSAHGTSTPTVDSAHGTSTPTV 711  
 OY 414 KEPTP-TTKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 472  
 DB 712 VSSAHGTSTPTVDSAHSTPTVDSAHSTPTVDSAHGTSTPTVDSAHGTSTPTV 764

OY 473 PTTKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 528  
 DB 765 HGTSTPTVDSAHSTPTST-PAADSSAHGTSTPTVDSAHST-PAADSSAHGTSTPT 817  
 OY 529 TP-KKPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 563  
 DB 818 TPTVDSAHSTPTVDSAHSTPTVDSAHSTPTVDSAHSTPTVDSAHSTPTVDSAH 876  
 OY 584 TTP-EKLAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 633  
 DB 877 TPTVDSAHSTPTVDSAHST-TPSTVDSAHSTPTVDSAHSTPTVDSAHSTPTVDSAH 935  
 OY 634 PKPEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 693  
 DB 936 PSTPADSS-PAHSTPTVDSAHST-PTPADSS-PAHSTPTVDSAHST-PTPADSS 963  
 OY 694 TTEPTSTSDKPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 752  
 DB 984 -PSTPADSSAHST-PAHSTPTVDSAHSTPTVDSAHSTPTVDSAHSTPTVDSAHSTPS 1033  
 OY 753 KPAPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 808  
 DB 1034 TP-PAADSSAHGTSTPTVDSAHSTPTVDSAHSTPTVDSAHSTPTVDSAHSTPT 1084  
 OY 809 PTTSEVSTPTTKEPTTTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPKEPAATTTTPK 867  
 DB 1085 DSSAH-STPTPAD-SSAHSTP-PTPADSSAHSTP-PTPADSSAHSTP-PTPADSSAH 1125  
 OY 868 MT 869  
 DB 1126 GT 1127

RESULT 10  
 CPN\_DROME  
 ID CPN\_DROME STANDARD; PRT; 865 AA.  
 AC 002910;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE CALPHOTIN.  
 GN CPN OR CAP.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-CANTON-S;  
 RX MEDLINE-93165730; PubMed-8434015;  
 RA Ballinger D.G., Xue N., Hashman K.D.;  
 RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds  
 RT calcium and contains a leucine zipper."  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).  
 CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO  
 CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL  
 CC OF CA+2 PER MOL OF PROTEIN.  
 CC -1- SUBUNIT: HOMODIMER (PROBABLE).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.  
 CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF  
 CC COMPOUND EYES AND OCELLI.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL  
 CC DEVELOPMENT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L02111; AAA28405.1; -  
 DR EMBL: L05080; AAA28420.1; -  
 DR PIR: A47282; A47282.  
 DR FlyBase: FBgn0010218; Cpn.

KW Calcium-binding.  
 FT CONFLICT 36 A -> AVAPAVVA (IN REF. 2).  
 FT CONFLICT 43 I -> T (IN REF. 2).  
 FT CONFLICT 64 I -> V (IN REF. 2).  
 FT CONFLICT 76 T -> A (IN REF. 2).  
 FT CONFLICT 100 P -> PE (IN REF. 2).  
 FT CONFLICT 126 VQ -> AP (IN REF. 2).  
 FT CONFLICT 154 I -> V (IN REF. 2).  
 FT CONFLICT 160 S -> T (IN REF. 2).  
 FT CONFLICT 534 A -> E (IN REF. 2).  
 FT CONFLICT 699 I -> T (IN REF. 2).  
 FT CONFLICT 703 V -> L (IN REF. 2).  
 FT CONFLICT 721 D -> E (IN REF. 2).  
 SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;

Query Match 7.28; Score 530; DB 1; Length 865;

Best Local Similarity 26.28; Pred. No. 1.3e-14;  
 Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;

OY 279 APTSKVLAKP-TPK---AETTKCPALTPPKKEPTTPPKKEASPTPKKEPTTKSAPTT 334  
 DB 8 SPVSAPVAAYVTSAAVAPQVVSAPVAP-----APAPVIAVTPVAPPTTLASVGPAT 61  
 OY 335 PKREPAPTTKSAPTTP-KEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 389  
 DB 62 VTIPAPAPIAASVTPVASVAPVVAAPTP-----PAASPTSTPVAQAQIPVAASAPVAP 116  
 OY 390 KEAPAPTPKKRPAP-----TTKKEBAPTTPKEBAPTTP-----TTKKEBAPTTPKEBAPTTP 435  
 DB 117 PVATPTPVVQIPVAPVIAVTPVAPVIAVTPVIAVTPVIAVTPVIAVTPVIAVTPVIAVTP 176  
 OY 436 PKBPA-PTAPKKRPAPTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 487  
 DB 177 AAVPAAPVAVVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 236  
 OY 488 PTTTKEP---APTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 543  
 DB 237 PLAAEFVVAAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 296  
 OY 544 TTP-----KEBAPTTPKKRPAPTPKEBAPTTP-----APTTPKKTLP 584  
 DB 297 ETPLAPVAVESOVAAVTVATPTTPAPPEPTTAPVVAETPEVASVAVETTPPVVPPV 356  
 OY 585 TPEKL-APTTPKEBAPTTPBELAPTPPEBPTTPPEB---APTTPKAAAPNTPKEBAPT 640  
 DB 357 AAESIPAPVAVATPVAPAT---LAVTBDVTAASAVBELPVIAVSPVSAVAETPVDLAP 413  
 OY 641 TPKEBAPTTPKEBAPTTPKETAPTP-KGAPVTLKE-----PAPTTPKKRPAPKE 689  
 DB 414 V---LPVAAEPVAPVAAETPETPAPASAPVIAVIAVIAVIAVIAVIAVIAVIAVIAVIA 469  
 OY 690 LAPTTKEPTSTSDKAPTPKGTAPTP-----KEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 741  
 DB 470 AAPVAVSTPP---TASVETTAAPVAVTEPIDVSVLSAALETPVAPVEVTTVAADV 527  
 OY 742 TLKEPA-----PTTPKKRPAPKELAPTTT---KGPTSTSDKRAPPT---PKETAPPT 787  
 DB 528 AAPAAADLIIEPPEPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 585  
 OY 788 PKREPAPTTKKRPAPTPPETPT 837

DB 586 EVAVAPITAPERIP-----EPEPSLATPTPEPIPVAPVIAQAVDAVEVPVETSTSIPT 641  
 OY 838 SAETPKALENSKPEPGVPPTTKPAATKPEMTTAKOKTERDRTTP--ETTAAAKMT 895  
 DB 642 TVE-FPVAVAKVLDPAI--TEAPVTOEDPDVAININGAPAPATEI-TTPVAEIVTAAAEVS 697  
 OY 896 KEVATTEKTESKITATTTTOVST-----TTODTTPFKITTLTKTTLA-PKVTTTKT 948  
 DB 698 DIAIPVLDPPVQELIAVAIEPDTKPAEIVVESTIPIEAPVPEVSKYAPVISEAPAA 757  
 OY 949 ITTTEINMKPEETAKPKDRATNSKATTPPKQPKPAKPKPTSTPKPTMPRVKPKTPT 1008  
 DB 758 EVPTIADNDPNTSVGISSEVPTIAEKPVVEPPTS--EIEQSSSPS--DSVPVAKITPL 813  
 OY 1009 PKRSTMPDELNPTSRIAEAM--LQTTTRNCPNSKLVVN-----PKS 1051  
 DB 814 LRDIQTTDVSLAIAATLDAIGERLKDQKARNOQVMDRLCEIEKILGPPKS 864

RESULT 11  
 PPL\_MYTCO  
 ID PPL\_MYTCO STANDARD; PRT; 872 AA.  
 AC Q25434;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCPPI).  
 GN PPL.  
 OS Mytilus coruscus (Sea mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OC NCBI\_Taxid=42192;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=FOOT.  
 RX MEDLINE=96394686; PubMed=8798340;  
 RA Inoue K., Takeuchi Y., Takeyama S., Yamaha E., Yamazaki F., Odo S.,  
 RA Harayama S.;  
 RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and  
 RT its evolutionary implications.";  
 RL J. Mol. Evol. 43:348-356(1996).  
 CC - FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
 CC - SUBCELLULAR LOCATION: SECRETED.  
 CC - TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
 CC - DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
 CC - PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND  
 CC ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D63777; BAA09850.1; -  
 DR InterPro: IPR002964; Adhesive\_plaq.  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PRO1216; ADHESIVEI.  
 DR PRINTS: PRO1217; PRICHEXTENS.  
 KW Signal; Repeat; Hydroxylation.  
 FT SIGNAL 1 20  
 FT CHAIN 21 872 POTENTIAL.  
 FT DOMAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.  
 FT DOMAIN 21 872 NONREPEPTIVE LINKER.  
 FT DOMAIN 124 872 TANDDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-  
 FT P-[PST]-[ST].  
 FT DOMAIN 184 192 NONAPEPTIDE 1.  
 FT DOMAIN 213 221 NONAPEPTIDE 2.  
 SQ SEQUENCE 872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;



```
FT REPEAT 432 443 3-4.
FT REPEAT 444 453 3-5.
FT REPEAT 454 460 3-6.
FT REPEAT 461 472 3-7.
FT REPEAT 473 479 3-8.
FT REPEAT 480 491 3-9.
FT REPEAT 492 498 3-10.
FT REPEAT 499 515 3-11.
FT REPEAT 516 522 3-12.
FT REPEAT 525 566 P-TYPE 4.
FT DOMAIN 572 613 P-TYPE 5.
FT DOMAIN 620 661 P-TYPE 6.
FT DISULFID 162 188 BY SIMILARITY.
FT DISULFID 172 187 BY SIMILARITY.
FT DISULFID 182 199 BY SIMILARITY.
FT DISULFID 307 333 BY SIMILARITY.
FT DISULFID 317 332 BY SIMILARITY.
FT DISULFID 327 344 BY SIMILARITY.
FT DISULFID 354 380 BY SIMILARITY.
FT DISULFID 364 379 BY SIMILARITY.
FT DISULFID 374 391 BY SIMILARITY.
FT DISULFID 526 552 BY SIMILARITY.
FT DISULFID 536 551 BY SIMILARITY.
FT DISULFID 546 563 BY SIMILARITY.
FT DISULFID 573 599 BY SIMILARITY.
FT DISULFID 583 598 BY SIMILARITY.
FT DISULFID 593 610 BY SIMILARITY.
FT DISULFID 621 647 BY SIMILARITY.
FT DISULFID 631 646 BY SIMILARITY.
FT DISULFID 641 658 BY SIMILARITY.
FT VARIANT 276 276 K -> E.
FT VARIANT 354 354 K -> R.
FT VARIANT 415 415 T -> A.
SQ SEQUENCE 662 AA; 67774 MM; F085277F1ED2FD40 CRC64;
```

Query Match 7.0%; Score 509.5; DB 1; Length 662;  
Best Local Similarity 28.1%; Pred. No. 6.6e-14;  
Matches 223; Conservative 48; Mismatches 276; Indels 247; Gaps 30;

```
QY 248 TTNKQSTGDEKETSKEIOSIKTSKDLAPTSKYLAKTPPAETTTKGPALTTRKPP 307
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 3 TTAATAVATGCTDTAAEGSAAAEKTA-----AGEVSAPPT--AAVAATGEDATT---- 51

QY 308 TPTTPKEBAPTKEPPTTKSAPTTKPEBAPTPTTKSAPTTKPEBAPTPTTKKBPAP 367
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 52 -----AAATAAEETTAAGEAPTPTTTTAPATAAGKAPPTAATAATPATAAGAPPTAT 103

QY 368 EPAPTTTKEBAPTPTTKS-APTTPKEBAPTTPPKKBPAPTTKPEBAPTPTTKPEBAP 425
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 104 GKAPATAAAPPVTAASKAPPTAAAGHSTAAAPPTTAASAKSKERSTSSSEEHCH 163

QY 426 --PTTKBPAP--TKPEBAPTAKKBPAPTPK-----EPAPTTKPEBAPTPTTKE 470
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 164 VKPSKREMGSGKITKQ-----CKKNKCCPEPKGHGHCPRPKGSHSEHHTTTTK- 218

QY 471 PSPPTTKEBAPTPTTKSAPTTKPEBAPTPTTKPEBAPTPTTKPEBAPTPTTKPEBAPT 530
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 219 -----APTIIQIATTTT--TPTTT-----TTTATKATPTT----- 245

QY 531 KKBPAPTTKPEBAPTPTTKKBPAPTPAPKPEBAPTPTTKETAAPTTKPKLTPTTPEKIA 590
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 246 -----TTTAKAPPTT-----TTTAKATTTT-----TPTT----- 270

QY 591 PTPPEKAPPTPEELAPTPPEEPPTTPEEBAPTTPKAAPNTKEBAPTTPKPEBAPTTP 650
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 271 -----TTTTKATPTPTTTTTP----- 289

QY 651 KEAPATTPKEBAPTTPKGTAPTTLKEBAPTTPKKAPAKELAPTTTKEPTSTSDKAPATT 710
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 290 -----TTTTKATTTTTTTSGECKMEPSK-----REDCGSGITESOCR 328

QY 711 PKG-----TAPPT--PKBAPTTPKEBAPTTPKGATPTTLKEBAPTTPPKKAPKELA 760
```

```
Db 329 TKGCCEPDSIPQTKMCFYTLTSQVADCKVEPQKRDCCGRGIT-----ADQCKKCCPDS 384
    || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 761 PTTTKGPTSTSDKBPAPTTKEBAPTTPKBPAPTTKBPAPTTETPEPTTSEVSTPTT 820
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 385 ISGTKWCFSYSTQVAA--TKTTTTPTTTPTTTPTTTTATTTTP-----TTTTTPTT 436

QY 821 KEPTTKSPDESIPPELSAEPPTKALENSPREKPEVPTTKPPAKPKPMTTAKDKTERD 880
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 437 TTTT-----TTKA-----TTTTPTTTPTTTT--KAT--- 463

QY 881 LRTTPTTAPKMTKETATTTTEKTSKITATTTQVSTTTODTPEFKITLTKTTLAP 940
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 464 -TTTTPTTTTP-----TTTTKAT-----TTTTPTTTPTTTTATTTPTTTT 510

QY 941 KVTTKKTIITTEIMNKPEEFAKPKDRATNSKATTPKQKTPKPKPTSKKKK-----T 996
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 511 TTTTAKATTTTSECKME-----PSKRADCGYPSITSOCKSCCPDSIPQTKMCFYS 566

QY 997 MPVRKPKPTPTTPR 1010
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 567 LPQVADCKVAPSSR 580
```

```
RESULT 13
RPB1_HUMAN STANDARD; PRT; 1970 AA.
ID RPB1_HUMAN
AC P24928;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).
GN POLR2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92178992; Pubmed=1542581;
RA Wintzerith M., Acker J., Vicaire S., Vigneron M., Keding C.;
RT "Complete sequence of the human RNA polymerase II largest subunit.";
RL Nucleic Acids Res. 20:910-910(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95347616; Pubmed=7622068;
RA Mita K., Tsuji H., Moriyasu M., Takahashi E., Neno M.,
RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
RT "The human gene encoding the largest subunit of RNA polymerase II.";
RL Gene 159:285-286(1995).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
RNA(N).
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PPM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

DR EMBL: X63564; CAA5125.1; -  
 DR EMBL: X74874; CAA52862.1; -  
 DR EMBL: X74873; CAA52862.1; JOINED.  
 DR EMBL: X74872; CAA52862.1; JOINED.  
 DR EMBL: X74871; CAA52862.1; JOINED.  
 DR EMBL: X74870; CAA52862.1; JOINED.  
 DR PIR: S21054; S21054.  
 DR MIM: 180660; -  
 DR InterPro: IPR000684; RNA\_pol.II\_repeat.  
 DR InterPro: IPR000722; RNA\_pol.A.  
 DR InterPro: IPR002879; RNA\_pol.A2.  
 DR Pfam: PF006623; RNA\_pol.A; 1.  
 DR Pfam: PF01854; RNA\_pol.A2; 1.  
 DR PROSITE: PS00115; RNA\_POL-II\_REPEAT; 43.  
 DR TRANSFERASE; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 DR DNA-binding; Nuclear protein; C2H2-type (POTENTIAL)  
 FT ZN FING 71 87 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 FT DOMAIN 1590 1958 W -> L (IN REF. 2).  
 FT CONFLICT 1067 1067 D -> Y (IN REF. 2).  
 FT CONFLICT 1449 1449  
 SQ SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;

Query Match 6.9%; Score 503.5; DB 1; Length 1970;  
 Best Local Similarity 33.3%; Pred. No. 2,8e-13;  
 Matches 161; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

QY 307 PRTTPEKPAATTPKE--PRTTTSKAPTTPKEPAATTTKSAPTTPKE----PAPTTPKE 360  
 DB 1507 PSMGSGSPMTNMGATATAYGAMSPSVSGMTGCAAGSPASASGSPSPYSAMS 1566  
 QY 361 PAPTTPKEPAATTTKEPAATTTKSAPTTPKEPAATTPKPAATTPKEPTT 418  
 DB 1567 PTPGSPGSPSPYIPSPGAMSPSPSPSPA-YEPRSPGTYTPQSPSPSPSPSP 1624  
 QY 419 TTPKEPAATTTKEPAATTTKPAATTPKPAATTPKEPAATTTKEPTTKE 478  
 DB 1625 TSSPSY-SPTSPNVSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-----SPTSPSY 1671  
 QY 479 PAPTTPKSAPTTTKEPAATTTKSAPTTPKEPSPPTTKEPAATTPKPAATTP 538  
 DB 1672 -SPTSPSPSPSPSY-SPTSPSPSPSPSY-SPTSPSY-SPTSPSY-SPTSPSP 1725  
 QY 539 KEPAATTPKPAATTTKPAATTTKPAATTPKETAATTPKTLPTTPEKLAATTPKPA 598  
 DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPNVT-PTSP-SYSPSPSP-SYSPSPNVT-T 1777  
 QY 599 PTTPPELAATTPPEPTTPEEPATTPKKAAPNTKEPAATTPKPAATTPKEP--APT 656  
 DB 1778 PTSP-NVSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1828  
 QY 657 TPKEAATTPPKGTAATTPKPAATTPK-KPAPELAATTTKEPTSTSDKPAATTPKGA 715  
 DB 1829 SP-SYSTSPSKY-----PTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1875  
 QY 716 PTTPKEPAATTPKEPAATTPKGAATTPKPAATTPKPAATTPKPAATTPKGA 774  
 DB 1876 PTSPPTV-SPTTPK-SPTSP--TYSPTSPVYTPSPK-----YSPSPSP--PTSPSPSP 1922  
 QY 775 PAPTTPKEAATTPKKEP--APTTPKPAATTPPEPTTSEVPTTKEPTTHKSPDE 832  
 DB 1923 YSPSPSP--TYSPTSPKSGTSPSPSY-SPTSPSP-----TYSPTSPSPAI-----SPDD 1965  
 QY 833 STPE 836  
 DB 1966 SDEE 1969  
 RESULT 14  
 RPBL\_MOUSE  
 ID RPBL\_MOUSE STANDARD; PRT; 1970 AA.  
 AC P08775;  
 DT 01-NOV-1988 (Rel. 09, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPBL).  
 GN POLR2A OR RPO2-1 OR RPII215.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-67280135; PubMed-3038894;  
 RA Ahearn J.M. Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.;  
 RT "Cloning and sequence analysis of the mouse genomic locus encoding  
 RT the largest subunit of RNA polymerase II.";  
 RL J. Biol. Chem. 262:10695-10705(1987).  
 RN [2]  
 RP SEQUENCE OF 1587-1970 FROM N.A.  
 RX MEDLINE-86068017; PubMed-2999785;  
 RA Corden J.L., Cadena D.L., Ahearn J.M. Jr., Dahmus M.E.;  
 RT "A unique structure at the carboxyl terminus of the largest subunit  
 RT of eukaryotic RNA polymerase II.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).  
 RN [3]  
 RP REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.  
 RX MEDLINE-92178992; PubMed-1542581;  
 RA Wintzerlth M., Acker J., Vicaire S., Vigneron M., Kedinger C.;  
 RT "Complete sequence of the human RNA polymerase II largest subunit";  
 RL Nucleic Acids Res. 20:910-910(1992).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MNNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: M12130; AAA40071.1; -  
 DR EMBL: M14101; AAA40071.1; JOINED.  
 DR PIR: A28490; A28490.  
 DR MGD: MGI:98086; Rpo2-1.  
 DR InterPro: IPR000684; RNA\_pol.II\_repeat.  
 DR InterPro: IPR000722; RNA\_pol.A.  
 DR InterPro: IPR002879; RNA\_pol.A2.  
 DR Pfam: PF006623; RNA\_pol.A; 1.  
 DR Pfam: PF01854; RNA\_pol.A2; 1.  
 DR PROSITE: PS00115; RNA\_POL-II\_REPEAT; 42.  
 DR TRANSFERASE; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 DR DNA-binding; Nuclear protein; C2H2-type (POTENTIAL)  
 FT ZN FING 71 87 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 FT DOMAIN 1590 1958 P -> R (IN REF. 1 AND 2).  
 FT CONFLICT 1498 1498 MISSING (IN REF. 1 AND 2).  
 FT CONFLICT 1499 1536  
 SQ SEQUENCE 1970 AA; 217175 MW; 7D76F38FD92A657E CRC64;

Query Match 6.8%; Score 498.5; DB 1; Length 1970;  
 Best Local Similarity 33.1%; Pred. No. 4,3e-13;  
 Matches 180; Conservative 101; Mismatches 168; Indels 95; Gaps 47;

Oy	307	PTPTTKEKPASTTPKE-PTPTTTSKAPLTTPKEPAPTTTTSKAPTTTKE-----PAPTTTKE	360
Db	1507	PSPMGGISPMATPMNOCATPAGVAMSGVSGMTPEAGCFSPSAAISDASGSPGSPAMWS	1566
Oy	361	PAPTTPEKAPATTTTKEKAPATTTTTSKAPTTPEKAPATTTTTPKKAPATTTTKEP-APTPTTKEPTP	418
Db	1567	PTPSPGSPGSPGSSYYTSPGAMSPSTSPSTSP-YPKSPGGYTTPQSPSTSPSTSPSY-SP	1624
Oy	419	TTTPEKAPATTTTKEKAPATTTTTPKAPATTTTTPKKEPAPTTTKEPAPTTTKESEPTTKE	478
Db	1625	TSPSPY-SPTSPBNYSTPSPSY-SPTSPSY-SPTSPSY-----SPTSPSY	1671
Oy	479	PAPTTTTSAPTTTKEKAPATTTTTSKAPTTTPKPSPTTTTKEKAPTTTPEKAPATTTTTPKKAPATTP	538
Db	1672	-SPTSPSYSPSTSPSY-SPTSPSYSPSTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY	1725
Oy	539	KEPAPTTPEKAPATTTTTPKAPATTTPEKAPATTTPEKAPATTTTPKTLTPTTPEKAPATTTPEKPA	598
Db	1726	SY-SPTSPSY-SPTSPSY-SPTSPBN-SPTSPMT-PTSP-SYSTSP-SYSTSPBN-T	1777
Oy	599	PTTPEELAPTTPEEPTPTTPEEAPATTTPKAAAPNTPEKAPATTTPEKAPATTTKEP-APT	656
Db	1778	PTSP-NVSPSTSPSY-SPTSPSY-SPTSP-SPTSPSDR-----YTQSP-TYTPSSPSYSPS	1828
Oy	657	TPKETAPTTPEKGTAPTLTKEKAPATTTK-KAPKELAPTTTKEPTSTTSKAPATTTPEKGT	715
Db	1829	SP-SYSTSPSPKYT-----PTSPSYSPSSSEYTPAPSK-YSPTSPSKYSTPTSPK-YS	1879
Oy	716	PTTPEKAPATTTPEKAPATTTPEKGTAPTLTKEKAPATTTTPKKAPKELAPTTTCKGPT-STTSDK	774
Db	1876	PTSPY-SPTTPK-SPTSP-SPTSP-SPTSPSYTPTTSK-----YSTS-PTYSPSTPK	1922
Oy	775	PAPTTPEKATPTTKEP-APTTPKKAPATTTPEPTTSEVSTPTTKEPTTHKSDE	832
Db	1923	YSPSTSP-YTSPSTSPKSGTSPSTSPGY-SPTSP-----TYSLTSPAI-----SPDD	1965
Oy	833	STPE	836
Db	1966	SDEE	1969
RESULT 15			
NFH_RAT			
ID	NFH_RAT	STANDARD:	PRT: 831 AA.
AC	P16884; Q63368;		
DT	01-AUG-1990 (Rel. 15, Created)		
DT	01-FEB-1994 (Rel. 28, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)		
DE	(NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).		
GN	NEFH OR NFH.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_Taxid=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RA	MEDLINE=89065087; PubMed=3143606;		
RA	Breen K.C., Robinson P.A., Wlon D., Anderson B.H.;		
RT	"Partial sequence of the rat heavy neurofilament polypeptide (NF-H).		
RT	Identification of putative phosphorylation sites."		
RL	FEBS Lett. 241:213-218(1988).		
RN	[2]		
RP	SEQUENCE OF 37-831 FROM N.A.		
RA	MEDLINE=88309090; PubMed=2457365;		
RA	Dattigny A., Pham-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,		
RT	Jolles P.;		
RT	"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and		
RT	in situ detection."		
RL	Biochem. Biophys. Res. Commun. 154:1099-1106(1988).		
RN	[3]		

RP SEQUENCE F 1-89 AND 243-313 FROM N.A.  
 RA MEDLINE=87080760; PubMed=2878828.  
 RA Robinson P.A., Wion D., Anderson B.H.;  
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide  
 RT (NF-H).";  
 RL FEBS Lett. 209:203-205(1986).  
 RN [4]  
 RP SEQUENCE OF 318-831 FROM N.A.  
 RA MEDLINE=89184647; PubMed=2928342;  
 RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,  
 RA Smulowitz M., Carroll Z., Emanuel B.S., Breitler J., Rubin L.;  
 RT "Cloning of a cDNA encoding the rat high molecular weight  
 RT neurofilament peptide (NF-H): developmental and tissue expression in  
 RT the rat, and mapping of its human homologue to chromosomes 1 and  
 RT 22.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).  
 CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
 CC -I- PM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SRINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -I- PM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENTLY WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -I- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 763  
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M37227; AAAA1693.1; ALT\_FRAME.  
 DR EMBL: X13804; CAA32038.1; ALT\_FRAME.  
 DR EMBL: M21964; AAAA1695.1; -  
 DR EMBL: J04517; AAAA1692.1; -  
 DR PIR: A30796; A30796.  
 DR PIR: A25649; A25649.  
 DR PIR: B25649; B25649.  
 DR PIR: S02003; S02003.  
 DR InterPro: IPR001664; IF.  
 DR Pfam: PF00038; filament; 1.  
 DR PROSITE: PS00226; IF.1.  
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.  
 FT NON\_TER . 1  
 FT DOMAIN 1  
 FT 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.  
 FT CONFLICT 164 164 L -> I (IN REF. 2).  
 FT CONFLICT 185 185 I -> S (IN REF. 2).  
 FT CONFLICT 193 193 L -> T (IN REF. 2).  
 FT CONFLICT 199 199 M -> T (IN REF. 2).  
 FT CONFLICT 346 346 K -> N (IN REF. 1).  
 FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).  
 FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).  
 FT CONFLICT 485 485 P -> S (IN REF. 2 AND 4).  
 FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).  
 FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).  
 FT CONFLICT 727 727 A -> V (IN REF. 4).  
 FT CONFLICT 757 759 AAP -> GST (IN REF. 4).  
 FT CONFLICT 769 769 T -> L (IN REF. 2).  
 FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).  
 SQ SEQUENCE 831 AA: 89486 MW: 180973C3f13BF768 CACG64;

Query Match

6.88; Score 498; DB 1; Length 831;



Best Local Similarity 27.2%; Pred. No. 2.3e-13;			
Matches 205; Conservative 63; Mismatches 320; Indels 166; Gaps 38;			
QY	146	KKVIESEEITEVDKDKKTKKPPVVDGAGSLDGFKTPDPTSTTQHNVST	205
Db	170	RKLLGEEC-----RIGFGPIP-----FSLTEGLPKI--PSMST--HIKVK	207
QY	206	SPKITTAKPINRPSLPNSDTSKETSIVNKTETTTNNKOTSDGKEKTSK	265
Db	208	EEKIKVVE-----KSEKETVIVBEQTEEQVTEVEEDKEAQGEEEAEE	255
QY	266	ETQSIEKTSKDLAPTSKVLAKTPKAETTKGPAITTKPEP-----TPITPKPEP-STTP	320
Db	256	GGEAAATTS-----PAAEAAASP-----EKETKSPVKEEAKEAASPAEAKSP	306
QY	321	KEPTPTTIKSAPTTPKPEP-----PTTTKSAPTTPKPEP-----APTTPKE	368
Db	307	AEVKSAAVAKSPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKS	365
QY	369	PA-----PTTTKEPA-----PTTTKSAPTTPKPEP-----PTTTKEPA---	409
Db	366	PAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKS	424
QY	410	-PTTPKEP-----TPITPKPEPAPTTKEP-----APTTPKEP-----PTAPKPA	456
Db	425	SPVTVKSPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKS	481
QY	457	PTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP	502
Db	482	GAKPPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKS	541
QY	503	PTTPKEPSPPTTTKEP-----PTTPKEPAPTTPKKAPTTTPKEPAPTTTPKEP	558
Db	542	PEKAKSPVKEEIKPPAEVKSPEKAKSPMRKEAKSPKAKTLDVKSPEAKPPAKKEAKRPA	601
QY	559	PTAPKEPAPTTPKETAPTTPKKLTPTTPKEKAPTTPEELAPTTPEEPTPTTP	618
Db	602	DIRSPEQVKSAPAEVKSPEKEETRT--EKVAPKEEVKSPV--EEV--KAKEPPKKVE	654
QY	619	EEPAPTPKAAAPNTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEP	678
Db	655	EETPATPKTEVKSDEAPKQAQRP-KAEKEPLETEKPKDSGGEAKKEA-----KEKK	709
QY	679	PTTPKKPAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKPEPAPTTPKPEPAPTTPKGT	738
Db	710	AAAPEEETPAKGVKEEAKPKKAEDAKA-----KEPSKPESEK-----KPK--	751
QY	739	APTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTPKETAPTTPKPEPAPTTPKK	798
Db	752	-----KEEVPAAPAEKKDKTE-----EKTTSKKREKPKMEAKAKED-----K	790
QY	799	PAPTTPEPTTSEVSTPTTTPKEPTTIHKSDE	832
Db	791	GLPQEPSKPKTERAKSSSTSDQKDSQSEKAPED	824

Search completed: April 26, 2002, 16:33:14  
Job time: 624 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:26:07 ; Search time 61.21 Seconds  
(without alignments)  
500.359 Million cell updates/sec

Title: AA7  
Perfect score: 7323  
Sequence: 1 MANKTLPYLLLLSVFVIO.....ARAITRSGQTLSKVVNCP 1361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep: \*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	7.1	805	US-09-103-429A-4	Sequence 4, Appl
2	510.5	7.0	1837	US-08-928-361B-5	Sequence 5, Appl
3	506.5	6.9	744	5202236-25	Patent No. 5202236
4	496.5	6.8	786	US-09-103-429A-3	Sequence 3, Appl
5	496.5	6.8	1721	US-08-700-651-5	Sequence 5, Appl
6	496.5	6.8	1721	US-08-928-361B-6	Sequence 5, Appl
7	488.5	6.7	826	US-07-638-431-2	Sequence 2, Appl
8	488.5	6.7	826	PCT-US92-00018-2	Sequence 2, Appl
9	476.5	6.5	652	5202236-13	Patent No. 5202236
10	455.5	6.2	960	US-09-219-849-5	Sequence 5, Appl
11	428	5.8	1185	US-09-041-886-23	Sequence 23, Appl
12	424.5	5.8	1867	US-08-479-537A-5	Sequence 5, Appl
13	424.5	5.8	1867	US-09-083-116-5	Sequence 5, Appl
14	424.5	5.8	2035	US-08-479-537A-2	Sequence 2, Appl
15	424.5	5.8	2035	US-09-083-116-2	Sequence 2, Appl
16	420.5	5.7	907	US-08-783-774-2	Sequence 2, Appl
17	420.5	5.7	907	PCT-US95-04611A-19	Sequence 19, Appl
18	419.5	5.7	2476	US-08-276-967-2	Sequence 2, Appl
19	417	5.7	829	US-08-642-255-132	Sequence 132, App
20	417	5.7	829	US-08-397-633A-53	Sequence 53, Appl
21	417	5.7	837	US-08-175-155-68	Sequence 68, Appl
22	417	5.7	837	US-08-477-509B-103	Sequence 103, App
23	417	5.7	837	US-08-642-255-101	Sequence 101, App
24	417	5.7	837	US-08-707-237A-75	Sequence 75, Appl
25	417	5.7	837	US-08-482-085B-103	Sequence 103, App
26	417	5.7	897	US-08-397-633A-50	Sequence 50, Appl
27	404.5	5.5	408	US-07-609-716-65	Sequence 65, Appl

28	404.5	5.5	408	4	US-08-475-411A-65	Sequence 65, Appl
29	404.5	5.5	408	4	US-08-478-029A-65	Sequence 65, Appl
30	398.5	5.4	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	5.4	682	1	US-08-397-633A-36	Sequence 36, Appl
32	385	5.3	1231	3	US-08-904-263A-4	Sequence 4, Appl
33	381.5	5.2	1537	1	US-08-325-267A-2	Sequence 2, Appl
34	381	5.2	1848	4	US-08-296-791-6	Sequence 6, Appl
35	381	5.2	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
36	379.5	5.2	1187	1	US-08-320-559-28	Sequence 28, Appl
37	379.5	5.2	1187	3	US-08-545-860D-28	Sequence 28, Appl
38	379.5	5.2	1187	3	PCT-US94-04496-28	Sequence 28, Appl
39	379.5	5.2	1210	1	US-08-320-559-26	Sequence 26, Appl
40	379.5	5.2	1210	3	US-08-545-860D-26	Sequence 26, Appl
41	379.5	5.2	1210	5	PCT-US94-04496-26	Sequence 26, Appl
42	376	5.1	960	4	US-09-219-849-6	Sequence 6, Appl
43	371	5.1	761	2	US-08-707-237A-84	Sequence 84, Appl
44	371	5.1	762	1	US-08-642-255-114	Sequence 114, App
45	371	5.1	762	1	US-08-397-633A-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1  
US-09-103-429A-4  
; Sequence 4, Application US/09103429A  
; Patent No. 6187558  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 No. 6187558th Tioga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,429A  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 805 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: peritrophic membrane  
; US-09-103-429A-4

Query Match 7.1%; Score 522; DB 4; Length 805;  
Best Local Similarity 27.8%; Pred. No. 3e-25;

```
Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;
QY 281 TSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEASTTPKPEPTTTIKSAP-----TTP 335
Db 9 TALGLVAARPEVSDAEKNLHHPDPXP--PAQXXLLPXEYDCTKFYCEYGLKFIAP 66
QY 336 KEPAPTT-----TKSAPTTKPEAPTTTKPEAPTTKPEAPTTKPEAPTTK 383
Db 67 RDCAPGTEFEKSAQTCVHAALAGCTLPGPAETT--QAPATTO--APTITQ--APTIT 119
QY 384 SAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 443
Db 120 QAPTTTQATTT-----QAPTTTQAPTTT-----QATTTQAPTTT----- 156
QY 444 PKKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 503
Db 157 --QAPTTTQ--APTITQ--APTITQ-----APTITQ--APTITQ--APTITQAA 198
QY 504 TTPKPEPTTKPEAPTTT--KEAPTTTTPKPKAPTTKPEAPTTKPEAPTTK 547
Db 199 TTPAATTTPAAT--TPAATTTPAATTTPAATTTPGVPAPTS---APVMPICELLNGCPADFD 253
QY 548 -----PA-----BPAPTTT 555
Db 254 IHLPHDKYCNLFYQCSNGYTFEQRCEGLYFNYPVORCDSPANVECDGEISPAVPTE 313
QY 556 K-----PA----- 558
Db 314 GNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERRCAGTHFSFELQ 373
QY 559 -----PTAP--KEAPTT-----TPKET 573
Db 374 QCDHIELVGCTLPGGESEEDVDDEDACTGWYCTEPIEWEPPLNGCPADFSIDHLLPHES 433
QY 574 -----APTTPKLPPTT-----PEKLAPTTPEKPA 599
Db 434 DCGYLOCVHGOIARPCGNLHFSNATQSCSEPVTAGCVFECDSNQCTSTAAPTAAP 493
QY 600 TTPPELAPTTPEPTTTPPEAPTTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTK 659
Db 494 TAAPTAAPTAAPTAAPSTVVVPPA--TPPATAPVPPTTAIPT-----PAPTAAPTAAPTA 548
QY 660 ETAPTT---PKGTAPTTKPEAPTTKPKAPKELAPTTTKEPTSTWSDKAPTTPKGTAP 716
Db 549 PESPTTVVPTTAAPTA---APT---AVPEIPTVTSAPTAAPT--AAPTAAAPTAAP 598
QY 717 TTPKPEAPTTKPEAPTTKPGTAPTTTLKPEAPTTKPKAPKELAPTTTGTST--TSDKP 775
Db 599 TTAVPEIPTT-----VTSPTTAAPT--AAPANT-----TVVPTTAAPTAAP 641
QY 776 APTTPKETAPTTKPEAPTTKPKAPTTTTPPTTSEVSTPTTKEPTTIHKSPDESTP 835
Db 642 AP---NTTVTAPTAAPT--AAPANTTVVPTTAAPTAAP-----PTVAH-----AP 685
QY 836 ELSAETPKALENSPKPEGVPTTKTAATKPE 867
Db 686 NTTAAP-----VTTTSAPATT--PE 703

RESULT 2
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
```

```
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1837 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-5
```

Query Match 7.0%; Score 510.5; DB 3; Length 1837;

Best Local Similarity 20.8%; Pred. No. 4.1e-24; Indels 559; Gaps 52;

Matches 314; Conservative 109; Mismatches 530;

```
QY 131 STTKRSPKPKPKKKT-----KKVIESEEITEVDKNNKRTKKKPTPKPPVVDKAGS 181
Db 174 SITSGELKDPNKQATISGSRGCKGKYSIDSSSTGRFVDSITGLPTDYPNCFNPVTGN 233
QY 182 GLDNGDFKVTTPDSTOHNKVVSTSPKITTAKP-----INPRSLPNSDTSKE--- 230
Db 234 LVSRSTGK--TIPNT---YAGVYRSNETKTPEPSANTNFFLLVDPKINAPCNSENSEFQV 288
QY 231 ----TSLVKNKETTVEKETTNTKQSTDDGKEKTTTSAKETOSIEKTSKDLAPTSKVL 285
Db 289 IFDMGSKVVIPTKCVGKHNTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTT 348
QY 286 AKPTPKAETTTKGPALTTKPEPTTPKPEASTTPKPEPTTTIKSAPTTKPEAPTTKS 345
Db 349 TTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 408
QY 346 APTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 405
Db 409 TTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 460
QY 406 KEPAPTTKPEPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 465
Db 461 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTAKKPT--TTTTTTTT 514
QY 466 TTKPESPPT----- 475
Db 515 TTTATTTTTTSEPSVIKPDENWCLEKNGCEAKGATYGVIGDKGRIENGMAFTWIPND 574
QY 476 ----- 475
Db 575 DTHVRFKVKDVGNTISVRCGKGAGKLEFPDRSLDFTIPPVAGHNSCSIIIVGVSQGGKI 634
QY 476 -----PKEPAP-----TTTKSAPTTTK 492
Db 635 HVSPYSGKDVSLISAPIQCELFNEFYCDCTAKYGAIHSGYQTSADFVTTTTAKFTTTT 694
```



```

896 KETATTTKTTESKITATTT 915
      : | | : | : | : |
720 YRAKPTYPSTYKAKPTYPST 739

DBD

RESULT 4
US-09-103-429A-3
; Sequence 3, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: cdna and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tioga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 786 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-3

Query Match 6.8%; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.18; Pred. No. 8.8e-24;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

QY 288 PTPKAETTTKGPALTTPKPEPTPTTKPEAPSTPTTKTSAPTTPKEPAPTTTKSAP 347
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 93 GPPPAE-TQAPATP---QAPTTTQAPTTTT-----QAPTTTQ---ATTTTQAP 135

QY 348 TTPKEPAPTTTKEPAPTTTPKPEAPTTTTSAPTTPKEPAPTTTPKKPAPTTPKE 407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 TTTQ--APTQQ--ATTQ---QAPTTTQ--APTQTQAPTITTQ--APTQQ--APTQQ- 183

QY 408 PAPTTPKPEPTPTTPKPEAPTTTKEPAPTTAPKPKAPPTTTPKEAPTT-PKEPA-- 464
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 -APTQTQAPTITTQAAATTPAATTTPAATTTPAA-TTPAATTTCGVAPTSPVWPPIC 241

QY 465 -----PTTTPKEPSPTPK- 477
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

```

Db 1153 NSINKDPVNTQYTSNTGNI-----INP--ETGKVIPIGSLPGLNYPSENTPQOTDEITGK 1206
QY 1079 ----MDYLRVFNQGIINP-----MLSDETNICNGK----- 1106
Db 1207 PVDVTYGLPYDSTGEIDPATKLPDGSVAGDEILTEVLNITTTDEVTLGTPIDLETGLPR 1266
QY 1107 -PVDGLTTLNRGNTLV 1120
Db 1267 DPVSGLPQLPNGTLV 1281

RESULT 6
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-6

Query Match 6.8%; Score 496.5; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 2.9e-23;
Matches 295; Conservative 86; Mismatches 495; Indels 459; Gaps 43;

QY 110 PTSPSSKKAPPSPGASOTIKSTTKRSPKPNKKTKTKVIESEITVKNKKNRTKKKP 169
Db 82 PTDPVSNCFNPVTC-NLVSRTGKTPN-----TVAGYVRNSETK--- 121
QY 170 TKPVPVDEAGSLDNGDFKVTPTDTSTQHNKYSTSPKITTAKPINPRPSLPNSDTSK 229
Db 122 TTPEPSANTYAGYVRN-ETKTTPEPSANT---NFLVDPKI-----NAPCNSENSF 167
QY 230 E-----TSLTVNKKETTVEETKTTTNNKQSTDGKEKTTSAKETQSIEKTS 274
Db 168 EGOQIFDMGSKVYIPYKCVGVKHTTTTTTTTTTTTTTTTTTTT----- 208

```

QY	1079	----	MDYLPVRYNOGLIINP-----	MLSDETNICMGK-----	1106
Db	1207	PVDVTGILPYDPSTGEIIDPATKLPICGSVAGDEILVELNITDVTGLPIDLETGLPLR	1266		
QY	1107	-PVDGLTTLRLNGTLV	1120		
Db	1267	DPVSGLPQLPNGTLV	1281		
RESULT	7				
US-07-638-431-2					
Sequence 2, Application US/07638431					
Patent No. 5198535					
GENERAL INFORMATION:					
APPLICANT: Hoffman, Stephen L.					
APPLICANT: Charoenvit, Yupin					
APPLICANT: Hedstrom, Richard					
APPLICANT: Khumsmith, Srisin					
APPLICANT: Rogers IV, William O.					
TITLE OF INVENTION: Protective malaria sporozoite surface protein					
TITLE OF INVENTION: immunogen and gene					
NUMBER OF SEQUENCES: 2					
CORRESPONDENCE ADDRESS:					
ADDRESSEE: A. David Spevack					
STREET: NMRDC Building 1 T-12 National Naval					
CITY: Bethesda					
STATE: MD					
COUNTRY: USA					
ZIP: 20814-5044					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Floppy disk					
COMPUTER: IBM PC compatible					
OPERATING SYSTEM: PC-DOS/MS-DOS					
SOFTWARE: Patent In Release #1.24					
CURRENT APPLICATION DATA:					
APPLICATION NUMBER: US/07/638,431					
FILING DATE: 19910110					
CLASSIFICATION: 424					
AUTORNEY/AGENT INFORMATION:					
NAME: Spevack, Avrom D.					
TELEPHONE: (301) 295-6759					
TELEFAX: (301) 295-4033					
INFORMATION FOR SEQ ID NO: 2:					
SEQUENCE CHARACTERISTICS:					
LENGTH: 826 amino acids					
TYPE: AMINO ACID					
TOPOLOGY: linear					
MOLECULE TYPE: protein					
US-07-638-431-2					
Query Match	6.7%	Score	488.5;	DB 1;	Length 826;
Best Local Similarity	26.4%	Pred. No.	3.9e-23;		
Matches	153;	Conservative	54;	Mismatches	223;
				Indels	149;
				Gaps	34;
QY	324	PTTTIKSAPTTPKEPA-PTTTKSAPTTPKEPA-PTTTKEP----	APTTPKEP-APTTPKE	376	
Db	273	TPCKVRCQCPQIPVPIPNKIPKSPNEEPVNDPNDPNDPNDPNDPNDPNDPNDPNDPNN	332		
QY	377	PAPTTPKSAPTTPKEP-APTTPKKP-APTTPKEP-APTTPKEP-APTTPKEP-APTTPKEP	432		
Db	333	PNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNN	392		
QY	433	PTTPKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT	490		
Db	393	RNPKRPNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKN	443		
QY	491	TKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE	548		
Db	444	PNKPNP----NEPSNPNKPNP----EPLNPNPNEPSNPNPNEPSNPNPNEPSNPNPNE	492		



QY 549 PATTITTKKAPAPKEP-APTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAP 607  
Db 493 PS-----NPNEPSNPNEPSNPNE--PSNPK-----PSNPNP--P 523  
QY 608 TPEEPTPTTPEAPATTAKAAPNTKPEP-APTTKPEPAPTTPEKPAPTTPEKPAPTT 666  
Db 524 SNPNP-----PLNPNP-----SNPNPNEPSNPNEP--PSNPK-----PSNPNP 564  
QY 667 KGTAPTTKAPAPKEPAPKAPKELAPTTTKEPTSTSDKPAPTTPEKPAPTTPEKPAPTT 726  
Db 565 -----PSNPNPNEP--PSNPK-----SNPEEPINPEELNPKPSNPNEESN 606  
QY 727 PKEPAPTTKGTAPTTKAPKAPKELAPTTTKEPTSTSDKPAPTTPEKPAPTTPEKPAPTT 786  
Db 607 KPEPI-----NPESNPKPEP-----INPDENENPLIODEPIERNDSNVPI 649  
QY 787 TKPEP-----APTTPKAPAPTTPEPTTSEVSTPTTKEPTTIHKSDESTPELSAE 840  
Db 650 LPIIPKGNNTPSNLPENPSSEVEYPRNDNGENSNNTMKSNNI---PNEPIPSPGDN 706  
QY 841 PTPKALENSPK-----EPGVPTTK 859  
Db 707 PYKGHEERIPKPHRSNDYVYDNNVNNKNDKEPEIPNNE 745

RESULT 8  
PCT-US92-00018-2  
; Sequence 2, Application PC/TUS9200018  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard  
; APPLICANT: Khushmith, Srisin  
; APPLICANT: Rogers IV, William O.  
; TITLE OF INVENTION: Protective malaria sporozoite surface protein  
; TITLE OF INVENTION: Immunogen and gene encoding  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: A. David Spevack  
; STREET: NWRDC Building 1 T-12 National Naval  
; STREET: Medical Center  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20814-5044  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00018  
; FILING DATE: 19920103  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, Avram D.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-4033  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 826 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US92-00018-2

Query Match 6.7%; Score 488.5; DB 5; Length 826;  
Best Local Similarity 26.4%; Pred. No. 3.9e-23;  
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;  
QY 324 TPTTIKSAPTTPKEPA-PTTTKSAPTTPKEPA-PTTTKEP-----APTTPKEP-APTITKE 376

Db 273 TPKVRCQPIPIPPVIPNKKIPEKPSNPNEPVPNDNDPNPNPNPNPNPNPNPNPNPN 332  
QY 377 PAPTITKSAPTTPKEP-APTTPKAP-APTTPKEP-APTTPKEP-APTTPKEP-APTTPKEP 432  
Db 333 PNNPN 392  
QY 433 PTTTKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 490  
Db 393 RRPKRNRNPKPKNKNPNKNPNKNPNKNPNKNPNKNPNKNPNKNPNKNPNKNPNKNPN 443  
QY 491 TKEPAPTITKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 548  
Db 444 PKNPNP-----NEPSNPKNPNP-----EPLNPNPNPNPNPNPNPNPNPNPNPNPN 492  
QY 549 PAPTITTKKAPAPKEP-APTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEKPA 607  
Db 493 PS-----NPNEPSNPNEPSNPNE--PSNPK-----PSNPNP--P 523  
QY 608 TPEEPTPTTPEAPATTAKAAPNTKPEP-APTTPKEPAPTTPEKPAPTTPEKPAPTTPE 666  
Db 524 SNPNP-----PLNPNP-----SNPNPNEPSNPNEP--PSNPK-----PSNPNP 564  
QY 667 KGTAPTTKAPAPKEPAPKAPKELAPTTTKEPTSTSDKPAPTTPEKPAPTTPEKPAPTT 726  
Db 565 -----PSNPNPNEP--PSNPK-----SNPEEPINPEELNPKPSNPNEESN 606  
QY 727 PKEPAPTTKGTAPTTKAPKAPKELAPTTTKEPTSTSDKPAPTTPEKPAPTTPEKPA 786  
Db 607 KPEPI-----NPESNPKPEP-----INPDENENPLIODEPIERNDSNVPI 649  
QY 787 TKPEP-----APTTPKAPAPTTPEPTTSEVSTPTTKEPTTIHKSDESTPELSAE 840  
Db 650 LPIIPKGNNTPSNLPENPSSEVEYPRNDNGENSNNTMKSNNI---PNEPIPSPGDN 706  
QY 841 PTPKALENSPK-----EPGVPTTK 859  
Db 707 PYKGHEERIPKPHRSNDYVYDNNVNNKNDKEPEIPNNE 745

RESULT 9  
5202236-13  
; Patent No. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
; PROTEIN  
; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,762  
; FILING DATE: 25-MAY-1990  
; APPLICATION NUMBER: 82,456  
; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945  
; FILING DATE: 24-NOV-1986  
; APPLICATION NUMBER: 650,128  
; FILING DATE: 13-SEP-1984  
; SEQ ID NO:13  
; LENGTH: 652  
5202236-13  
Query Match 6.5%; Score 476.5; DB 6; Length 652;  
Best Local Similarity 29.7%; Pred. No. 1.7e-22;  
Matches 221; Conservative 71; Mismatches 284; Indels 169; Gaps 45;  
QY 207 PKITTAKPINRPSLPPNSDTSKETSLSVTNKVETTKETTTNKQTSDECKEKTSAKE 266  
Db 11 PKMTVPPTYKPKPSYPP-----TYKSKPTY-----KPKIT----- 40  
QY 267 TQSEKTSKADLAPTSKVLAKAPTPKAETTTKGPALTTPKEPTTPKEP-----PASTTPKE 322  
Db 41 -----YPPTYK--AKPS-----YPPTYKPKKTYPPTYKPKLTYPPTYKPKP 79

OTHER INFORMATION: amino acid sequence  
US-09-219-849-5

Query Match 6.2%; Score 455.5; DB 4; Length 960;  
Best Local Similarity 23.0%; Pred. No. 5.5e-21;  
Matches 195; Conservative 64; Mismatches 338; Indels 251; Gaps 46;

QY 110 TSPSSKKAPPSPGASQITSTTKRSKPPNKKTKKVIIESEITEVDKNNKTKKKP 169  
DB 23 PAGPPGSRD-PGPPGAPGAPGPPGSRDPCPGF-----AP 55  
QY 170 TPK-PPVVDAGSLDNGDFKVTTEDTSTQHNKIVSTSPKITTAKPINRPSLPSNDS 228  
DB 56 PGAGPPGSRDPC-----PPGAPGAPGPPG-----80  
QY 229 KETSLTVNKETTVETTTNNKQISTDCKEKTTSAKETQSIKTSKADLAPTSKVLAKP 288  
DB 81 -----RDPGPP-----GAP 89  
QY 289 TPKAETTKGPAITTPKEPTPTTPKEPASTTPKEPTPTTPKEPASTTPKEPASTTP 348  
DB 90 GPAGPPGSRDPC-PGAPGAPGAPGPPGSRDPCGAPG-PAGPPGSRDPCGAPG-----PG 138  
QY 349 TPKEPAPTTPKEPAPTTPKEPA-PTTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP- 405  
DB 139 APGAPGPPGSRDPCGPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 198  
QY 406 -KEPAPTTPKEPAPTTPKEPAPTTPKEPA-----440  
DB 199 SRDPGP-PGAPGAPG-----PGSRDPCGPPGAPGAPGAPGAPGAPGAPGAPGAPG 252  
QY 441 -PTAPK-KPAPTTP-----KEPAPTTPKEPA-PTTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 494  
DB 253 GPAGPPGSRDPCGPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 306  
QY 495 APPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTP-----KKPAPTTPKEPAPTTP-----KE 548  
DB 307 GP-----PGAPGAPGPPGSRDPCG-PGAPGAPGPPGSRDPCG-PGAPGAPGPPGSRD 356  
QY 549 PAPPTTKKPA-----PTAPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 600  
DB 357 PGPPGAPGAPGPPGSRDPCGPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 412  
QY 601 TPELAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP-----KEPAPTTPKEPA-- 654  
DB 413 APGAPGPPGSRDPCG-PGAPGAPGPPGSRDPCGPPGAPGAPGPPGPPGPPGPPGPPGPPG 470  
QY 655 -----PTTPKET-APTTPKET-----APTTLKEPAPTTPKPP-APKELAPTTPKEPT 699  
DB 471 KGAPGAPGAPGPPGSRDPCGPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 530  
QY 700 STTSKAPAPTTPKGA-PTTPKEPAPTTP-----KEPAPTTPKGA-PTTLKEPAPTTPKPP 754  
DB 531 PPGA--PGAPGPPGSRDPCGPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 586  
QY 755 AP-----KELAPTTPKETST--TSDKAPAPTTPKEPAPTTPKEPAPTTPKPPAPTTP-- 804  
DB 587 GPAGPPGSRDPCGPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 644  
QY 805 --ETPPPTTSVSTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 857  
DB 645 SRDPGPPGAPGAPGPPGSRDPCG-PGAPGAPGPPGSRDPCGPPGAPGAPGPPGPPGPPGPPG 703  
QY 858 TKTPAATK 865  
DB 704 AHGPAGPK 711

RESULT 11  
US-09-041-886-23  
; Sequence 23, Application US/09041886  
; Patent No. 6235872

QY 323 PTPTTIKSAPT-TPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTP 379  
DB 80 SYPTTKSKTTPKPTTPPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPT 138  
QY 380 TTTKSAPTTPKEPAPTTPKSAPT-TPKEPAPTTP-----PKEPAPTTPKEPAPTTPKEPAPT 434  
DB 139 PTYKPKSYB--PSYKTKKTYPTTPKLTTPPTTPKPSYPTTPKPSYPTTPKPSYPTTP 196  
QY 435 TPKEPAPTAPKKA-PTTPKEPA-PTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTP 485  
DB 197 YP-----PTYKAKSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPT 252  
QY 486 SAPTTTKEPA-----PTTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 532  
DB 253 AKPTYKAKPTYKAKPTYKAKPTYK-----PTYKAKSYPTTPKSAKSYPTTPKSAKSYPT 308  
QY 533 PA--PTTPKEPA--PTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 588  
DB 309 PSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPT 362  
QY 589 LAPITPE--KPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 643  
DB 363 AKPTYKAKPTYKAKPTYKAKPTYK-----PTYKAKSYPTTPKSAKSYPTTPKSAKSYPT 413  
QY 644 EPA-PTT-----PKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 696  
DB 414 KPTVPSTYKAKPSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPT 473  
QY 697 EPTSTSDKAPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 754  
DB 474 KPTVPSTYKAKPTYK--AKPSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPTTP 521  
QY 755 APKELAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 807  
DB 522 TVK-AKPTYKAKPSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPTTPKSAKSYPT 576  
QY 808 PPTTSEVSTPTTPKEPTTHKSPDESTPELSAETTPKALENSKEPVGPTTKTPAATKPE 867  
DB 577 PTYKAKPSYPTTP-----YKAKPSYPTTPKSAKPTYKA-----KPTVPST---YKAKPS 620  
QY 868 MTTAKDKTTERDLRTTPETTTAAP 892  
DB 621 YPPTHKAKPTYKAKPSYPTTHKAKP 645

RESULT 10  
US-09-219-849-5  
; Sequence 5, Application US/09219849  
; Patent No. 6150081  
GENERAL INFORMATION:  
; APPLICANT: VAN HEERDE, GEORGE V.  
; APPLICANT: VAN RIJN, ALEXIS C.  
; APPLICANT: BOUWSTRA, JAN B.  
; APPLICANT: DE WOLF, FREDERIK A.  
; APPLICANT: MOOBROEK, ANDREAS  
; APPLICANT: WERTEN, MARC W.T.  
; APPLICANT: WIND, RICHEL D.  
; APPLICANT: VAN DEN BOSCH, TANJA J.  
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN  
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE  
; FILE REFERENCE: 2728-2  
; CURRENT FILING DATE: 1998-12-23  
; CURRENT APPLICATION NUMBER: US/09/219,849  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 960  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative

GENERAL INFORMATION:  
 APPLICANT: Bredesen, Dale E.  
 TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
 TITLE OF INVENTION: Polypeptides and Methods of Use  
 NUMBER OF SEQUENCES: 72  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP  
 STREET: 4370 La Jolla Village Drive, Suite 700  
 CITY: San Diego  
 STATE: California  
 COUNTRY: United States  
 ZIP: 92122

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/041.886  
 FILING DATE:

CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P-LJ 2626  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1185 amino acids

TYPE: amino acid  
 TOPOLOGY: linear

MOLECULE TYPE: protein  
 US-09-041-886-23

Query Match 5.8%; Score 428; DB 4; Length 1185;  
 Best Local Similarity 22.1%; Pred. No. 3.7e-19;  
 Matches 237; Conservative 114; Mismatches 429; Indels 292; Gaps 50;

QY 117 KKAPPP-----SGASOTIKSTTKRSKPPNKKTKKVVIESEITEVKNKNRKKKP 169  
 Db 18 KEAPGPREELSRASFGVSSSSDCKAEKSRQAKAKARVEEASTPKYKQGRSEE-- 75  
 QY 170 TPKPPVVDKAGSLNDGDFKVTTPDTSTQHNKVSPTSKITAKPINRPSLPPNSDTSK 229  
 Db 76 -----ISESES-----BETNAPKTKTEQELPRQSPSDLDLSD 109  
 QY 230 ETSLVNKEITVETKETTINKOTSTDGKEKTSKETSKEKTSKADLAPTSKVLAKPT 289  
 Db 110 GRSL--NDGSSPRDIDQNRSTS-----PSIYSPGSVENDS----- 145  
 QY 290 PKAETTKGPAITPKPTTP--KEPASTTKPTTPTTIKSAPTTPKEPATTTKSAP 347  
 Db 146 DSSSGLSQGA--RPYHPPLFPSPQPDSTPRQEAFF-----EPHPSVTPTGY 194  
 QY 348 TTPKEPATTTKPEATTTPKEPATTTPKAPT--TTKSAPTPK----- 390  
 Db 195 HAPWEP-PTSRMFQAP--PGAPPPHPLYPGGTGVLSPGPMGPKGGGAASSVGGNGCK 251  
 QY 391 -EPATTPPKPAPTTPKEPATTTPKEPTTTPKEPATTTPKEP-----TTPKEPAPT- 443  
 Db 252 QHPPTTPISSVSSGASGAPTKP-----PTTPVGGNLPSPAPPPANFHVTPNLPPPAL 307  
 QY 444 -PKPAPTTPKEPATTTPKEPATTTPKESPTTPKE-----PAPTTPKSAPTTPKE-- 494  
 Db 308 RPLNNASAP----PGLGAQPLPGHL--PSYVAMGGMGLPPGP--EKGTLPSPHS 358  
 QY 495 APTTPKSAPTTP-KEPSPTTPKEPATTTPKEPATTTPKAPTTPKAPTTPKE-PAPT 552  
 Db 359 LPPASSAPAPWRRPYSS 416

QY 553 TTKKPAPTAPKEPATTTPKETAPTTPKKLT-----PTTPE--KLAPTTPEKAPTTPPEEL 605  
 Db 417 SL-----SVSNQP-----PKYTQPSLPSQAVWSQGGPPPPYGRLLANSNAHPGFPSTG 467  
 QY 606 APTTPEEPTPT-----TPEEPAP-----TTPKAA 629  
 Db 468 AQSTAHPEVSTHHHHHQQOQQOQQOQQOQQOHHGNSGPPPCAFPHPLEGSSHHAHYAM 527  
 QY 630 APN-----TPKEPAPTTPKEP-----APTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAP 679  
 Db 528 SPSGLSLRPPTPGPAHLPPPHSHQSVSYQAGPNGPVPVSSSSSSSSSTSGQSVPCS--HPSP 585  
 QY 680 TTPKKPAPKELAP---TTTKEPTSTTSDDKPAATTPKGTAPTTPKEPAPTTPKEPAPTTPK 736  
 Db 586 SQGQGAPYPPFPVPTVTTSATILSTVIATVASSPAGYKTASPPGPPPYKRAPSPGAYK 645  
 QY 737 GTAPTTLKEPAPTTPKAPKAPKELAPTTPKGTSTTSDDKPAATTPKGTAPTTPKEPAPTTP 796  
 Db 646 TATPPGYKPGSP-----PSFTGTGPGYRG---TSPAGPGTFFKPGSPV--GPGPLPP 694  
 QY 797 KKPAPTTPETPPPTTSVSTPTTKEPTTIHKSPDE--STPELSAETPKALENS--PKE 852  
 Db 695 AGPS-GLPSLPPPPAPASOPPLS--ATQIKQEPAEYETPE---SPVPPARSPSPPKV 748  
 QY 853 PGVPTTKTAAKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTEKT-TESKIT 911  
 Db 749 VDPFHASQSAFENKHLDRGNSCARSDLYFVP---LEGSKLAKRADLVEKVRRAEOR 805  
 QY 912 ATTQVSTTTQDTPFKITTLTKTTLAPKVTTTKTITTTIMNKPDEETAKPKDRATNS 971  
 Db 806 AR-----EKEKEREREREK 820  
 QY 972 KATTPKQKPTKAPKKTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLO 1031  
 Db 821 ERERERELERSVKLAQEGRAP-----VECPSLGVPVPHR-----PPFEGSAVA----- 865  
 QY 1032 TTPRPQNTNSKLVEVNPKSEDAAGAEGETPHMLRPHVFMPEVTPDMDYLP 1083  
 Db 866 -TVPPYLGPDTPALRT--LSEYA-----RPHVMSPGNRNHPFVYP 902

## RESULT 12

US-08-479-537A-5  
 ; Sequence 5, Application US/08479537A  
 ; Patent No. 5861381  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHAMBERN, Pierre  
 ; APPLICANT: KIEN, Marie-Paule  
 ; APPLICANT: LATHE, Richard  
 ; APPLICANT: HAREUVENI, Mara  
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
 ; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
 ; STREET: P.O. Box 1404  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: United States  
 ; ZIP: 22313-1404  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/479,537A  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: FR 90/13101  
 ; FILING DATE: 23-OCT-1990

[illegible]

	2331	Conserved	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2433	2434	2435	2436	2437	2438	2439	2440	2441	2442	2443	2444	2445	2446	2447	2448	2449	2450	2451	2452	2453	2454	2455	2456	2457	2458	2459	2460	2461	2462	2463	2464	2465	2466	2467	2468	2469	2470	2471	2472	2473	2474	2475	2476	2477	2478	2479	2480	2481	2482	2483	2484	2485	2486	2487	2488	2489	2490	2491	2492	2493	2494	2495	2496	2497	2498	2499	2500	2501	2502	2503	2504	2505	2506	2507	2508	2509	2510	2511	2512	2513	2514	2515	2516	2517	2518	2519	2520	2521	2522	2523	2524	2525	2526	2527	2528	2529	2530	2531	2532	2533	2534	2535	2536	2537	2538	2539	2540	2541	2542	2543	2544	2545	2546	2547	2548	2549	2550	2551	2552	2553	2554	2555	2556	2557	2558	2559	2560	2561	2562	2563	2564	2565	2566	2567	2568	2569	2570	2571	2572	2573	2574	2575	2576	2577	2578	2579	2580	2581	2582	2583	2584	2585	2586	2587	2588	2589	2590	2591	2592	2593	2594	2595	2596	2597	2598	2599	2600	2601	2602	2603	2604	2605	2606	2607	2608	2609	2610	2611	2612	2613	2614	2615	2616	2617	2618	2619	2620	2621	2622	2623	2624	2625	2626	2627	2628	2629	2630	2631	2632	2633	2634	2635	2636	2637	2638	2639	2640	2641	2642	2643	2644	2645	2646	2647	2648	2649	2650	2651	2652	2653	2654	2655	2656	2657	2658	2659	2660	2661	2662	2663	2664	2665	2666	2667	2668	2669	2670	2671	2672	2673	2674	2675	2676	2677	2678	2679	2680	2681	2682	2683	2684	2685	2686	2687	2688	2689	2690	2691	2692	2693	2694	2695	2696	2697	2698	2699	2700	2701	2702	2703	2704	2705	2706	2707	2708	2709	2710	2711	2712	2713	2714	2715	2716	2717	2718	2719	2720	2721	2722	2723	2724	2725	2726	2727	2728	2729	2730	2731	2732	2733	2734	2735	2736	2737	2738	2739	2740	2741	2742	2743	2744	2745	2746	2747	2748	2749	2750	2751	2752	2753	2754	2755	2756	2757	2758	2759	2760	2761	2762	2763	2764	2765	2766	2767	2768	2769	2770	2771	2772	2773	2774	2775	2776	2777	2778	2779	2780	2781	2782	278
--	------	-----------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/083,116

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/479,537

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1867 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1727

OTHER INFORMATION: /note= "The amino acids spanning

128 to 1727 constitute a repeated

region wherein the repeat

is 20 amino acids, 17 of which are fixed.

The number of such

repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa

which is the codon for Pro or Ala wherein

Pro = CCT, CCC, CCA

or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa

which is the codon for Thr or Asn wherein

Thr = ACT, ACC, ACA

or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa

which is the codon for Pro or Ala wherein

Pro = CCT, CCC, CCA

or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note= "Amino acids 1 to 21 are a

21 amino acid precursor sequence."

OTHER INFORMATION: US-09-083-116-5

Query Match 5.8%; Score 424.5; DB 4; Length 1867;  
Best Local Similarity 23.7%; Pred. No. 1e-18;  
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 231 TSLTVNKETTIVETKETTITNKOTSTDKKEKTTSAKETQSIKTSKAD-LAPTSKVLAKPT 289  
Db 16 TVLTV-----VTGSHASSTPGGEKETSATORSSVPSSTKNAVSMSTSSVLSSHS 65  
QY 290 P-KAETTTKGP--ALTTTPKEP-----TPTTKEPASTTTPKEPTTIKSA 332  
Db 66 PGSGSSTTQGDVTLAPATEPASGSAATGQDVTSVPTPALGSTTTPAHADV---SAP 122  
QY 333 TTPKEPAPTTT-----KSAPTTTPKEP-----APTTPKEPAPTTTPKEPAPT 373  
Db 123 --DNKPAPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 180  
QY 374 TKE--PAPTTT-----KSAPTTTPKEPAPTTTPK---KPAPTTTPKEPAPTTTPKEPTPTT 420  
Db 181 APDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 240  
QY 421 PKE--PAPTTKEP-----APTTPKEPAPTTAPK---KPAPTTTPKEPAPTTTPKEPAPT 467  
Db 241 APDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 300  
QY 468 TKEPSPTTPKEPAPTT--TKSAPTTTTPKEPAPTT-----TKSAPTTTPKEPSPTTTPK---E 516  
Db 301 APDXRP--XPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 359  
QY 517 PAPTTTPKEPAPTTTPK---KPAPTTTPKEPAPTTTPK---EPAPTTTTPKEPAPTTTPKEPAPT 568  
Db 360 SAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 419  
QY 569 TPKETAP-----TTPKKLTPTTTPKEPAPTTTPKEPAPTTPEELAPTTPEEPTPTTPEEPAPT 624  
Db 420 SAPDXRPPGSTAPXAHGVT-----APDXRPPXPGSTAPXAHGVTAPDXRPPXPGSTAPXAHGVT 471  
QY 625 TPKA---AAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 679  
Db 472 APXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPXPGSTAPXAHGVTAPDXRPPXPGSTAPXAHGVT 530  
QY 680 TTP-----KKPAPKELAPTTTKEPTSTTSDKPAP---TTPKGTAPTTTPKEPAPTTTP 727  
Db 531 TAPXAHGVTAPDXRPPXPGSTAP--XAHGVTAPDXRPPXPGSTAPXAHGVT-----APDXR 585  
QY 728 KEAPTTTPKG---TAPTTLKEPAPTT-----KKPAPKELAPTTTKEPTSTTSD 773  
Db 586 PXPGSTAPXAHGVTAPDXRPPXPGSTAPXAHGVTAPDXRPPXPGSTAP--XAHGVTAPDX 644  
QY 774 KPAP--TTPK---ETAPTTTPKEPAPTTTP-----KKPAPTTTPETPTTTPSEVSTP 817  
Db 645 RPXPGSTAPXAHGVTAPDXRPPXPGSTAPXAHGVTAPDXRPPXPGSTAPXAHGVTAP 702  
QY 818 TTTKEPTT-----LHKSPDESTPELSAEPKALENSPKBEGVPTTKTPAA-----863  
Db 703 DXRPPXPGSTAPXAHGVTAPDXRPPXPGSTAPXAHGVTAPDXRPPXPGSTAPXAHGVTAP 762  
QY 864 -TKPEMTTAKDK---TTERDLRTTPTTAAAPKMTKETATTEKTESKITATTQVTS 919  
Db 763 DXRPPXPGSTAPXAHGVTAPDXRPPXPGST--APXAHGVTAPDXRPPXPGSTAPXAHGVT 820  
QY 920 TTTQDTPPKITTLKTTTLAPKVTTTTKTITTTTEIMNKPEETAAPKDRATNSKATTPKQ 979  
Db 821 APDXRPPXPG-----STAPXAHGVT-----SAPDXRPPXPGSTAPXAHGVTAPDXRPPX- 868  
QY 980 KPTAKPKKSTTKPKTTPRVRKPKTTPTPRKMSTMPNELNPTSRIAEAMLOTTPTRPQ 1039  
Db 869 -GSTAPXAHGVTAPDXRPP--XPGSTAPXAHGVTSA-PDXRPP-----PGST 911  
QY 1040 -PNSKLVEVNPKSEDAAGAEGETPMLLRPHVFMPEVTPDMDYLPVPVQGIIPMLSD 1098  
Db 912 APXAHGVTAPDXRPPXPGSTAPXAH-----GVTSAPDXRPPXPGSTA-----PXAHG 957









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:49 ; Search time 114.61 Seconds  
(without alignments)  
871.345 Million cell updates/sec

Title: AA8

Perfect score: 7064

Sequence: 1 MAWKTLPIYLLLSVEVIO.....ARATTRSGQTLSKVWVNCNP 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*

1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1078.5	15.3	3020	2 A43932	mucin 2 precursor,
2	950	13.4	1664	2 T18262	S-layer protein -
3	855	12.1	1489	2 T31108	cyst germination s
4	802	11.4	1274	2 T16251	hypothetical prote
5	798.5	11.3	1367	1 S48478	glucan 1,4-alpha-g
6	784	11.1	1188	2 S49915	extensin-like prot
7	776.5	11.0	2187	2 T30826	nascent polypeptid
8	672	9.5	1344	1 A35175	mucin 1 precursor,
9	671	9.5	1151	2 T18535	high molecular mas
10	662	9.4	1229	2 T35697	hypothetical prote
11	661	9.4	3507	2 T34513	hypothetical prote
12	646.5	9.2	3570	2 T45025	hypothetical prote
13	635.5	9.0	7962	2 T38346	mucin MUC5B, trach
14	633	9.0	489	2 T11622	elastic titin - hu
15	632	8.9	761	2 C84672	extensin class 1 p
16	630.5	8.9	990	2 T1618	hypothetical prote
17	626.5	8.9	971	2 T19431	nucleolar phosphop
18	625	8.8	6642	2 T29757	hypothetical prote
19	622.5	8.8	839	2 F75518	protein UNC-89 - C
20	608	8.6	801	2 T29018	hypothetical prote
21	607.5	8.6	924	2 S27923	gene LF3 protein -
22	605	8.6	379	2 S50125	larval glue protei
23	559.5	7.9	350	2 S22456	hydroxyproline-ric
24	556.5	7.9	856	2 T16543	hypothetical prote
25	556.5	7.9	875	2 S23760	polyphenolic adhes
26	555.5	7.9	1630	2 A53577	ascites sialoglyco
27	554	7.8	2232	2 T34434	hypothetical prote
28	551	7.8	620	2 S06733	hydroxyproline-ric
29	545	7.7	873	2 A47283	calphotin - fruit

30	542	7.7	369	2 S20500	hydroxyproline-ric
31	538.5	7.6	416	2 J00465	extensin precursor
32	533	7.5	1087	1 QFMSH	neurofilament trip
33	532	7.5	756	2 T27642	hypothetical prote
34	530.5	7.5	1162	2 JH0557	exo-alpha-sialidas
35	530	7.5	865	2 A47282	calcium-binding pr
36	522	7.4	328	2 J00985	hydroxyproline-ric
37	522	7.4	1459	2 T32271	hypothetical prote
38	518	7.3	813	2 S70795	vsA protein precu
39	518	7.3	866	2 T45462	membrane glycoprot
40	518	7.3	1072	1 A37221	neurofilament trip
41	512.5	7.3	867	2 T45463	membrane glycoprot
42	509.5	7.2	662	2 A45155	mucin FIM-C.1 - Af
43	508	7.2	1832	2 T31113	mucin-like glycopr
44	505	7.1	700	2 A54641	interspersed repea
45	504.5	7.1	606	2 A43427	neurofilament trip

#### ALIGNMENTS

RESULT 1

A43932

mucin 2 precursor, intestinal - human (fragments)

N:Alternate names: mucin SMUC-41

C:Species: Homo sapiens (man)

C>Date: 10-Mar-1993 #sequence\_revision 12-Apr-1996 #text\_change 05-Nov-1999

C:Accession: A49963; A45106; B45106; B33532; A61257; P00328; P00329

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t

A:Reference number: A49963; MUID:94132002

A:Accession: A49963.

A:Molecule type: mRNA

A:Residues: 1-639 <GU1>

A:Cross-references: GB:L21998

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up

A:Reference number: A45106; MUID:93016075

A:Accession: A45106

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 626-1895 <GU2>

A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396

A>Note: sequence extracted from NCBI backbone (NCBIP:116706)

A:Accession: B45106

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 2037-3020 <GU3>

A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398

A:Experimental source: colon

A>Note: sequence extracted from NCBI backbone (NCBIP:116698)

R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,

J. Clin. Invest. 88, 1005-1013, 1991

A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym

A:Reference number: A43932; MUID:91358717

A:Accession: A43932

A:Molecule type: DNA

A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>

A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864

A>Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)

R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989

A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl

A:Reference number: A33532; MUID:89197956

A:Accession: B33532

A:Molecule type: mRNA

A:Residues: 1916-2193 <GU4>

A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874

A:Experimental source: intestine

R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991  
A:Title: Human bronchus and intestine express the same mucin gene.

A:Reference number: A61257; MUID:91086481

A:Accession: A61257

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>

A:Experimental source: bronchus

R: Xu, G.; Huan, L.; Khatiri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner,

Biochem. Biophys. Res. Commun. 183, 821-828, 1992

A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-

A:Reference number: PQ0328; MUID:92198477

A:Accession: PQ0328

A:Molecule type: mRNA

A:Residues: 2328-2468 <XUG>

A:Cross-references: GB:M86523

A:Experimental source: small intestine

A:Accession: PQ0329

A:Molecule type: protein

A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>

C:Genetics:

A:Gene: GDB:MUC2

A:Cross-references: GDB:120203; OMIM:158370

A:Map position: 11p15.5-11p15.5

C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von

C:Keywords: glycoprotein; intestine; tandem repeat

F: 2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 15.3%; Score 1078.5; DB 2; Length 3020;

Best Local Similarity 28.1%; Pred. No. 9.1e-39;

Matches 378; Conservative 88; Mismatches 478; Indels 399; Gaps 52;

QY 43 TCNCDY--NCQHYMECCDFKRVCTAELSC-----KQRC-----FE---76

DB 1079 SCSCDRGGDCECSAVASQAQCEKAGCVFWRTPDLCPIFCDYNNPHECEWHYEPGC 1138

QY 77 -SFERGECID-----CDAQKKYDKCQDVFESFCAEVKDNKKNKTKK---118

DB 1139 NRSFECRTINGIHSNISVYLSGCPYRCPK-DR--PIVE-----EDLKKCVTADKCGC 1189

QY 119 ---PTKPP-----VV--DEAGSLD---NGDF---KVTTPDTST 147

DB 1190 YVEDTHYPGASVPTBEETCKSCVCTNSSQVVCPRPEGKILNQDGAFCYWEICGNGTV 1249

QY 148 TOH---NKVSTSPKITTAKINRPSLPNSDTSKETSIVNKETTVETKET--TTNK- 201

DB 1250 EKHFNCISITRPSLTFTTILPTTP-----TSFTTTTTTTTPTSSVLTTPKL 1301

QY 202 ---QSTDGKERTTS--AKETQSIEKTSKD-----227

DB 1302 CCLWSOWINEDHPSSGSDGDRPFDGVCGAPEDIERSVKDPHLSLEQHKGKVCQDVSV 1361

QY 228 -----LAPTSKVLAKPTPKAETTKGPAALT 253

DB 1362 GFICKNEDQFGNGPFGICDYKIRVNCWPMDCITTPSPPTTPSPPTTTTLPPTTT 1421

QY 254 KPEPTTPKPEASTPKPTTIKSAK--TTPKEAPTTTKSAPTPKPEAPTTKEPA 312

DB 1422 PSPPTTTTTPPTTPSPPTTTTTPPLPTTPSPPLSTTTTTPPTTPSP-PTTTPSP 1480

QY 313 PTPKPEAPTTTKEPAPTTKSAK--TTPKPEAPTTKPKAPTTKPEAPTTKEPPTTP 371

DB 1481 TTPSPPT--TTTTPPTTTPSPMTTTPPASTTTLPTTTPSPPTTTTTPPTTTP 1539

QY 372 KEPAPTTKEPAPTTKPEAPTTAPKKAPTTKPEAPTTKPEAPTTKESPTTKPEAP 431

DB 1540 SPPTTTPITB-PTSTTLPTTPPTTPSPPTTTTTPPTTPSPPTTTPSPPTTTTTP 1598

QY 432 TTPKSAK--TTPKPEAPTTTSAK--TTPKESPTTTPKPEAPTTKPEAPTTKPEAPTTK 489

DB 1599 TTPSPPTTTTTPPTTTPSPPTTTPPTTTLPTTTPPTTTPSPPTTTPPTTTP 1658

QY 490 EPAPTTKPEAPTTTTKKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 549

DB 1659 PPTTTPSPPTTTPPTTTPPTTTPSSPTTTP--SPPTT---TWTTPS--PTTTPSSPT 1710

QY 550 TTPPELAPTTTPEPTTTPPTTTPPEAPTTPKAAAPNTPKPEAPTTPKPEAPTTPK 608

DB 1711 TTTTPSSSTTP--SPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSSPTTTP 1769

QY 609 ----KETATTP-----KG-----618

DB 1770 FSPFSTTTTTPCPLCNWGLDSCPNHFKPGGDELIGDVGCGWGAANISCRATMP 1829

QY 619 -----TAPTTTKEPA 628

DB 1830 DVPIGOLGOTVVCDVSVGLICKNEDQKPGVIPMAFLNVEINVQCCVQPTTM--T 1886

QY 629 PTPPKKPAKELAPTT--TKEPTSTSDKP-----APT--TPKGTAPTTKPEAPTTKPEA 681

DB 1887 TTTTENPTPTTITTTTTPPTTTPSTQSPNGLOAPTPTPISTTTTTPPTTPTGTQPT 1946

QY 682 PTPKGTAPTTLKEAPTTKPKPAKELAPTTTTPKGTST--TSDKPAKPAKPAKPAK 740

DB 1947 -TPTTITTTTTPPTPTGTQPTTTLITTTTTPPTPTPTSTKSTVTPITTTTAT 2005

QY 741 PAPTTPKPK-----APTPEPTTTPPTTSEVSTPTTK-----EPTTIHKSPDESPELS 788

DB 2006 PPTGTQPTTPTPISTTTTTPPTTPTTGTSTGPTTSTAPIAELTTSNPPSPSTPQS 2065

QY 789 APTPKALENSKPEKGPVPTTKPAATKPEMTTAKDKTTERDLRTTPTTAAKPKTKET 848

DB 2066 RTSSTPLTSTLLTLPALPAEMTSTAPPTTSGHFLSPSPPTTSPPTTPR 2125

QY 849 ATTETTESKITATTTQVSTTTQDTPTFKITT---LKTTLAPKVTITTKTITTEIM 905

DB 2126 GTTT--GSSAPTPSVQITTSANTPTPLSTPLIIRTLGLRPVSSVLICCLNDY 2183

QY 906 NKPEE-----TAKPDRAINSKATTPKPKOKPKAPK 936

DB 2184 YAPGEEVINGTYGDTCYFVNCSLCTLEFVNWSCPSTPPTPSK--STPTPSKPSSTPS 2242

QY 937 KPTSTKPKTKMPVRPKTTPTRKMTSTMPELNPTSRATAEAMLOTTTRNPTNSKLVE 996

DB 2243 KPTPGTKPCEPDPDP--QENETWJLDCDFM--ATCKYNNTVEIVKVE 2288

QY 997 VNP-----KSEDAGG-----AEGETPHMLL-----RPHV 1020

DB 2289 CEPPTMPTCSNGLOPVYRVEDPDGCGWHWCDCYCTGWD--PHVTFDGLYYSQGNCTYV 2347

QY 1021 FMPEVTPDMDYLRPNVNOGIIIN 1043

DB 2348 LVEEISPSVD-----NFGVIID 2364

RESULT 2

T18262

S-layer protein - Clostridium thermocellum

C:Species: Clostridium thermocellum

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T18262

R:Fujino, T.; Beguin, P.; Aubert, J.P.

J. Bacteriol. 175, 1891-1899, 1993

A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellul

e.

A:Reference number: Z18847; MUID:93209931

A:Accession: T18262

A:Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: DNA

A:Residues: 1-1664 <FU>

A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 13.4%; Score 950; DB 2; Length 1664;

Best Local Similarity 31.0%; Pred. No. 1.6e-33;





QY 636 APKELAPTTTKEPTTSDKPAPTTPKGTAPTTTKEKPAP-----TTPKEPAPT-TPKGTAPT 691  
 Db 730 TESSAPVPTPS-SSITTESSAPVPTSTTTESSAPVPTSSSTTESSAPVPTPSSSSTE 788  
 QY 692 TLKCPAPTTKKPAPKELAPTTTKGPTSTSDKPAPTTTPKETAPTTKEPAP-----TTP 746  
 Db 789 SSSAPVPTPSSSTTESSAVPPTPSSSNITSSAPSSITPSSSTESSVVPVPTPSSSSTE 848  
 QY 747 KKPAPTTTPPTTSEVSTPT-----TTKEPTTIHKSDESTPELSAEPTEKALENSPK 801  
 Db 849 SSSAPVSSSTTESSAVPPTPSSSNITSSAPSSI--PFSSTTESFNGT-TVTPSSSK 904  
 QY 802 EPGVPTTKTAAAKPEMTTAKDKTTERDLRTPTTETAAPKWKETAATTEKTTESKIT 861  
 Db 905 YPGSQTETSVSSTETIVPTKTTTSTVPTSTTITTTVCSTGTSNAGETTSGCSPKTVT 964  
 QY 862 AT--TTQVSTTTTQDTPPKITTKTTLTTLAPKVT---TKKTIITT-EIMKKPETAKPK 915  
 Db 965 TTVPTTTTSTVTTSSITTTTVCSTGTSNAGETTSGCSPKITTTTVCSTSPSETA--- 1021  
 QY 916 DRATNSKATPKQKP-----TKAPKKPTSTKKPKTTPRVRKPKTTPT 958  
 Db 1022 -----SESTTPTTPTTVTVTVTVTVEYSTKPGGEITTFVTKNIPITLYTTIAPT 1076  
 QY 959 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRP 1018  
 Db 1077 P--SVTTVNTFTTT-ITTTVCSTGT-----NSAGETTSCSPKTVTTTTPC 1120  
 QY 1019 HVPMPEVTPDMDLPRVPNOGIIINPLSDETNICNGKXPVGLTTLRNGTLVAFRGHYFW 1078  
 Db 1121 STGTGYTTEATLVTAVTTVTTVTTSTESGTGNSA-GKTTTGYTTKSVPPT-----YVT 1173  
 QY 1079 MLSPFPSPSPARRITEVWGIPSPIDTVFTRC---NCEGKT 1115  
 Db 1174 TLAPSAPVTPATN----AVPTTIT--TECSAATNAAGET 1207

RESULT 6  
 S49915  
 C:Species: Zea mays (maize)  
 C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
 C:Accession: S49915  
 R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.  
 submitted to the EMBL Data Library, June 1994  
 A:Description: Pex genes: pollen-specific genes with extensin-like domains.  
 A:Reference number: S49915  
 A:Accession: S49915  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1188 <RUB>  
 A:Cross-references: EMBL:Z34465; NID:9600117; PIDN:CAA84230.1; PID:g600118

Query Match 11.1%; Score 784; DB 2; Length 1188;  
 Best Local Similarity 29.9%; pred. No. 1.3e-26;  
 Matches 230; Conservative 59; Mismatches 359; Indels. 122; Gaps

QY 119 PTPKPPVVDEAGSLONGDFKVTTPTDSTTQHKNKVTSPKITTAKPINRPSSL-----PPN 174  
 Db 460 PTPHSPPAD-----DYVPPTPPVPGKSPPTSPSQVQPPAASTPPPSLVKLSPPQ 510  
 QY 175 SDTSKETSILTVNKETTVETKETTITNKQTSDDGKEKTTSAKETQSTEKTSAKDLAPTSKV 234  
 Db 511 APVG-----GPPPVKTTSPAPIG-----SPSPPPVPVSV 541  
 QY 235 -----LAKPTPKAETTTKGPALTTPKPTPTTTPKEPASTTPKPTPTTIKSAPTTPKEPA 289  
 Db 542 SPPPPVKSPPPAPVGSPPPEKSPPPAPVAPSPPPVKSP--PPPTLVASPPPPVKSP 599  
 QY 290 PTTTKSAPTTP--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 346

```
Db 600 PPAVASPPPPVKSPPPPPTVVASPPPPAVASSPPPPMKSGPPPPPTVSSPPPPPEKSPPPPP 659
Qy 347 PKPAPPTTKEPAPTT-----PKE---PPTTTPKEPAPTTTKEPAPTTTKEPAPTTAPKKP 397
Db 660 PAKSTPPPEYPPPTSVKSSPPPEKSLPPTTLIPSPPPQEKPTPSTPSKP-PSSPEKP 718
Qy 398 APPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 457
Db 719 SP--PKEPVSSPQT-----KSSPPAPVSSPPPTVSSPPALAPVSSPPVSKSSP---- 768
Qy 458 EPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 517
Db 769 PPAPLSSPPAPQVKSPPPPVQVSSP--PPAPKSSPPPLAPVSSPPQVEKTSPPPPALSSPP 827
Qy 518 TTPKETAP---TTP---KKLTPTTPEKLAPTTTKEPAP---PTSTTSDKPAPTTTPKGTAPTT 564
Db 828 LAPKSSPHVVSSPPVVKSSPPPPAPVSSPPLTPKPPASPPPAHVSSPPVVKSTP--PA 885
Qy 565 PTTPEEPAPTTKAAAPNTKE-PAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 623
Db 886 PTTVISP--PSEPKSSPPTPVSLPPIVKSSPPPPAMVSSP-PMTPKSSPPPVVSSPPTT 943
Qy 624 LKE---PA-----PTTPKK---PAPKELAPTTTKE---PTSTTSDKPAPTTTPKGTAPTT 669
Db 944 VKSSPPAPVSSPPATPKSSPPPPAPVNLPPPEVKSSPPTTPVSSPPPA---PKSSPPPPAP 1000
Qy 670 -KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 728
Db 1001 MSSPPPEVKSSPPPPVVKSSPPPPVVKSSPPPPVVKSSPPPPVVKSSPPPPVVKSSPP 1057
Qy 729 -TPKETAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 787
Db 1058 KSPPPAPLSSPPPPVKSPPPPAPVSS--SPPPVKSPPPPAPVSSPPPPPKSPPPPAPV 1115
Qy 788 SAEPTPKALENSPKGPGVPTTKTTPAATKEMTTTAKDKTTERDLRTTTPET 837
Db 1116 SPPAP-----VKPPSLP-PPAPVSSPPPVVTPAPPKKEQSLPPPAES 1158

RESULT 7
T30826
nascnt polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphanAC into a muscle
A:Reference number: Z20889; MUID:96312450
A:Accession: T30826
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: EMBL:U48363; NID:gl666688; PID:gl666689; PIDN:AA818732.1
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphanAC into a tissue-specific DNA-binding activ
C;Keywords: alternative splicing; DNA binding; transcription factor

Query Match 11.0%; Score 776.5; DB 2; Length 2187;
Best Local Similarity 27.6%; Pred. No. 5e-26;
Matches 290; Conservative 119; Mismatches 416; Indels 227; Gaps 48;

Qy 98 PDYESCAEYKONKRTKKTPKPPVVDKAGSLDNGDFKVTTPDTSTTQHNK----- 152
Db 809 PKVDPIMSDVTPTSPKTKSATATPK-----DTATLSLKSVPAY 847
Qy 153 VSTSPKITTAKPINRPSLPNSDTSKETSL--TVNKET-----TVETKETTTNKQTS 204
```

```
Db 848 TSLSPKAPVAPSNEATIVPTETPTSLKNLAAATPKETLAISIPKVTSPSPQKTPKSVS 907
Qy 205 TDGKEKTTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGPALTTTTPKEPTTTPKE 264
Db 908 LKAGPAMTSKKAT---EIAASKDVSPQ--FPKEVPLLOHV---PPTSPKSPVSDTLSG 959
Qy 265 PASTTTPKEPTPTTIKSAPTTKE-PAPTTTTSAPITTPK-----EPAPTTTKEPAPT 314
Db 960 ALTSPKPGPPATLAETPTPKKSPKAASKKTATPSPGVTAVPLEIPPCSKKAPKA 1019
Qy 315 TPKEPAPTTTKEPAPTT--TKSAPT-----TPKEPAPTTPKK-P 350
Db 1020 APKESSATSSSRKRAPKTAVSKGVTAVPLEISLKETSKSATCEKSSASPSP 1079
Qy 351 APTTKEPAPTTTKEPAPTTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 407
Db 1080 KTAGKE---TPGGVTAVPPEISLPPKETPQONATPNSLAASSQKRSPTKVPKRETPG 1136
Qy 408 ---TTPKEPAPTTTKEPSTTPK-----EPAPTTTTSAPITTPK-----EPAPTTTKEPAPT 457
Db 1137 GVTAMPLEIPSAPQAPKTAVPKQIPTPEDAVTILAGSPLSPKASKATAAPKEAPATFSV 1196
Qy 458 -----EPSPTTKEPAPTTTKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTP 496
Db 1197 GVIAVSGEISPSPKTSKTAAPKENSATLPPKRSKTAAPKETPATSSGVTAVPSEIS 1256
Qy 497 KEAP-----TTTKKPAATAPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTP 548
Db 1257 SPPTPASKGVPTVLTTPKGNALAE--SPASPKKVPKTAAPETSTTP-----SPKIPK 1309
Qy 549 PTTPEELAPTTPEPTPTT-----PBEAPTTTTPKAAAPNTTPKEPAPTTTTPKE-PAPIT 599
Db 1310 VAGPKESATTPSKKTKTAVPKETSAPSEGVAVPLEIPSPRKPAPKTAAPKEIPAPS- 1368
Qy 600 PKPAPTTTKEPAPTTTTPKGTAPTTTTPKEPAPTTT-----KPPAPKE 639
Db 1369 -PGAITAPVQIIPSPKSGSKAGSKE-TPTTSPSEGVTAAPLEIPISKKTSKMASPKE 1426
Qy 640 LAPTTTKEPTS-----TTSDEKAPTTTTPKGTAPTTTTPKEPAPTTTTPKE-PAPITPKGTAPT 692
Db 1427 TLVTPSSKLSQTVGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLTPSPK-DAPIT 1485
Qy 693 LKEPAPTTPKKAPKELAPTTTTPKGTSTTSDKAPTTTTPKETAPTTPKEPAPTTTTPKAPAT 752
Db 1486 LAE-SPSSPK-APKTAAPPSER-VITVPEKPA-TPQASGTTASKVPVPAETQEVAVS 1541
Qy 753 TPETP-----PPTSEVSTPTTK-----EPTTIHKSDESTPELSAETPKALENSPK 801
Db 1542 SRETPTVPVAVPPVKNPSSHKKTSKTIKELKAPATLPPSPTKSPKIPSSKKAPRT--SAPK 1599
Qy 802 E-PGVPITTKPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTTKEKTESKI 860
Db 1600 EFPASPSIK-PVTT--SLAQTAAPSLOKAPSTIPKENLAAPAV---LPVSSKSPAAPA 1652
Qy 861 TATTQTSTTTQDPTTKITLTKTTLAPKVTTTKTITTTIMNKPEETAPKDRATN 920
Db 1653 RASASLPATAAPQATPKAETATIPSCKAAATETPIETSTAPSELEGAPKETSE-----TSV 1708
Qy 921 SKATTPPKPKTKAPKPKPTSTKPKTMPRVKPKTTTPRKMSTMPDELNPTSRABAML 980
Db 1709 SKVLMSPP-----PKASSSKRASTLP-----ATTLSLKEASVLS---- 1744
Qy 981 QTTTRPNOTPNSKLVEVNPKSEDAAGAGETP 1012
Db 1745 -----PATSSGKDSHISPVSS-DACSTGTTP 1770

RESULT 8
A35175
mucin 1 precursor, repetitive splice form A [validated] - human
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episia
ncretic mucin; polymorphic epithelial mucin (PEM)
N:Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor
```



Db 601 AP-PAHGVTSAPDTRPAGSTAPPAHGVTS-----APDTRPAGSTAPPAHGVTSAPDTRP 655  
QY 695 EPAPTTT-----KKPAPKELAPTTTKGPTSTSDKAP--TPK-----ETAPTTT 738  
Db 656 APGSTAPPAHGVTSAPDTRPAGSTAP-PAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 714  
QY 739 KEPAPTTT-----KKPAPTTTPTTPTTSEVSTPTTKEPTT-----IHSPD 781  
Db 715 PAPGSTAPPAHGVTSAPDTRPAGS--TAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPD 772  
QY 782 ESTPELSAETPKALENSPKPEGVPTTKTAA-----TKPEMTTAKDK-----TTERD 830  
Db 773 TRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPD 832  
QY 831 LRTTPTT-----TAAPKMTKETAATTEKTTESKITATTTQVITSTTODTTT--FKIT 881  
Db 833 TRPAGSTAPPAHGVTSAP-----DTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGV 888  
QY 882 TLKTTTTLAPKVTTTKKTTT--EIMNKPETAKPKDRATNSKATTPKPKQTKAPKKPT 939  
Db 889 SAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAG--GSTAPPAH 946  
QY 940 STKKPKTMPRVKTKTTPRKMTSTMPELN-----TSRIAEAMLOT--TTRP---NOTPN 991  
Db 947 VTSAPDTRP---APGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPP 1003  
QY 992 SKLVEVNPKSDAGAEGETPHMLLRPHVFMPEVTPMDYLPRVPMOGIILNPMLSDET 1051  
Db 1004 AHGVTSAPDTRPAGSTAPPAH-----GVTSAPDNRPALGSTA-----PPVHNVS 1049  
QY 1052 ICNGKPVGLTLRLNGTLVAFRGHYFWMLSPFSPS 1087  
Db 1050 ASGSAGSASTLVHNGTSARATTTTPASKSTPFSIPS 1085

RESULT 9  
T18535  
high molecular mass nuclear antigen - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18535  
R:Shimada, K.; Harata, M.; Mizuno, S.  
J. Cell Sci. 110, 3031-3041, 1997  
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick  
A:Reference number: Z18955; MUID:9803440  
A:Accession: T18535  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1151 <SHI>  
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

Query Match 9.5%; Score 671; DB 2; Length 1151;  
Best Local Similarity 25.9%; Pred. No. 8.6e-22;  
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

QY 120 TPKP-----PVVDAGSLGNDGFKVTTPTDSTTQHNKVS-----PKITTAKPINRPS 170  
Db 113 TTPPSOGPAGTPPSOGAAGAPKGDGTAQPSGTSKAGCKPAQDVPRATTA-ATEARP- 170  
QY 171 LPPNSDTSKETSITLVNKKETTTETTKN-----QTSDGKETTSAKETQSTIEKTSADL 228  
Db 171 -----ASAASPTVPKATAEATVTAASQSAPKAATDAAVTAA---SQSAPKATV-EV 219  
QY 229 APTSKVLAKTPKAETTTKGPALTTPKEPTT-TTPKEPASTTPKEP-----TPTTTKSAPT 283  
Db 220 KPAATAVAKEAKAVTAAAAAPKATAEAKPAPVTSPTIPCSSAEAKPLTAASFTASKA--T 277  
QY 284 TPKEPAPTT-----TKSAPTTTPKEPAPTTTK----- 309  
Db 278 AEAKVPATASLMATKATAEAKPAPSPVPKATTTDKAVTATAPKAGPDVKPAVAVCABA 337  
QY 310 EPAPTTTPKEPAPTTTPKEPAPTTTKSAPTT-----PKPA-----PTTPKKPAPTTTPK 356

Db 338 KPAPPPPOQLPKAAAAAAPTGTCLKPATAPPHGSPRANSHVTVTTPNPNVRAAAATVP- 396  
QY 357 EPAPTTTPKEPPTT-----PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 412  
Db 397 -TAGAVPKRSTGTTTAAAPQOPVP---KAAPVTPPSPOQAVPRAATAAA-----APVTPQ 448  
QY 413 P---APTTPKESPTTPKEPAPTTTKSAPTTTKEPAP---TTTKSAPTTTPKEPPTTTKE 466  
Db 449 PVTKAATTNATPPPOIPKKAATTTATVTPPOQIPKAGTDAAPPPAVPRAPSDGRAAT 508  
QY 467 P-----APTTPKEPAPTTTPKEPAPTTTPKEPAP---TTPKEPAPT-TTKPAPAP 512  
Db 509 PGVNAATDPOKPPPTPQSPVSAVTEPKPQRAAPPPSNEATPAVPSPNLKSPLPTIP 568  
QY 513 K-----EPAPTTK---ETAPTTPKLTPTTPSKLAPTTTPKEPAPTTPELA----- 556  
Db 569 KVPVLMALTPOQVTAQMTQLAATKPSIPVKASPK-ALMTPPPPPPGLPRALAAKLLG 627  
QY 557 -PTTP-----EETPTTTP---EEPAPTTT-----KAAAPNTPK----- 586  
Db 628 LPSSPVASAMHAKVTPRPLPASVPMAASPASILGPDARVALATNAASPGAKPEAAGNG 687  
QY 587 ---PAPTTTPKEPAPTTTPKEPAPTTTPKETAPT-----TPKGTAPT----- 622  
Db 688 TLMAPMAAANTOMAPIGAAGAAQTAPMGAAHTHVSPMGAGGATQMSPTGAANTHMSPIGA 747  
QY 623 -----TLKEPAPTTTPKPKAPKELAPTTTKEP-----TSTTSDKAPAPTT 661  
Db 748 GGATQSPMGAANTQSPMGCAATTTQSPMGAAATQVTSAGNTMQVSP 807  
QY 662 KG--TAPTTTPKEPAPTTTPKEPAPTTTPKGAPTTLKEP---APTTPKPKAPKELAPTTKGP 717  
Db 808 MGAATPPQTPSVGAATTP-QPSPM-----GAATTLMSPMGAATTPQ---ESPMAVTTQPP 859  
QY 718 ---TSTTSDKP-APTTPKET---APTTPKEP---APTTPKEP-APTTPETP----- 758  
Db 860 PMAATNTQPPMAASTPOSTPMGAATTTQSPMGATTTQSPMGASTPQAPPTVAGSPT 919  
QY 759 PTTSEVSTPTTPKEPTTIHKSDESTPEL--SAEPTPKALENSPKPEPGVPTTKTTPAATKP 816  
Db 920 PPPPIPPSPTAOTSPOPMKSPPPDPKAPSAATQSPAHHVANASPGV-TAVSPA---P 975  
QY 817 EMTTAKDKTTERDLRTTPTTITTAAPKMT-KETATTTETKTESKITATTTQVSTTTTQDT 875  
Db 976 IGVTEASPSADGARLSPGTAATDGPKASPAATADVTEAATD--VTAATAVPA----EA 1029  
QY 876 TPKITTLTKTTTLAPKVTTTTKTITTTIMNKPEETAKPKDRATNSKATTPKPKQKPTKAP 935  
Db 1030 AP-----TKAKRSSSPASPA--- 1077  
QY 936 KKPTSTKKPKTMPRVKPKTTPTRKMTSTMPELNPTSRITAEAMLQTTTTPNQTTPNSKLV 995  
Db 1078 -----PAVGDGOQOMTPGAASQSVPP-----VTEAAVQ----- 1104  
QY 996 EVNPKSEDAGGAGE 1010  
Db 1105 EAAAAAAGAGERE 1119

RESULT 10  
T25697  
hypothetical protein F16F9.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T25697  
R:Fulton, B.  
submitted to the EMBL Data Library, August 1996  
A:Description: The sequence of C. elegans cosmid F16F9.  
A:Reference number: Z20071  
A:Accession: T25697  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA



QY	73	RCFESFERGECDDCAQCKKVDKCCPDYVESCAEVKONKNRTKKKTPPKPPVVDEAGSG	132
Db	34	RFKVIKHIKKNNCTCSC-----KCPVD-----APSNP-----	61
QY	133	LONGDFKYVTTDPTSTQHKNKYSTSPKITTAKPINPRSLPPNSDTSKETSUTYKNKETT--	190
Db	62	-----FDVSTTISSINNDN-VDIGPS-----GDSNPTGSGWQEIATVGGGVTKS	106
QY	191	-----VETKETTINKOTS---TDGKEKTTSA-----KETQSIEKTS---A	225
Db	107	EHINIDSSVEVEKKVYTSDDASTNAPTTGKSDSTTPEIITGIWVINSKSESVDSTMSTRS	166
QY	226	KDLAPTSKVLAKTPKCAETTTKGPALTTKPEKPTTTPKEPASVTPKPEPTTIKSPATPP	285
Db	167	TTLSPTELLNSP---ETLVSDDSSSTSTQTSFQDNTTEIAS--PMETNNITT--EATTS	218
QY	286	KEPAPTTKS-----APPTPKCAPPTTKPE---APTPKPE---PAPTTTKPEPA	328
Db	219	VEPSVSTLASDETTVTAIAESTTTVIAEVSTTTTTEPTTTAESTTKSTTKKAPATTEET	278
QY	329	PTTTSKAPPTTKPEAPTTPKKPA---PTTP---KEPAPTTKPEPT---PTTPKPEAPTT	378
Db	279	PTTTTEEVTTTAEASTTTSSETSTEKPTTPLINKIAGATGATKPEPTHFPVGTGTPNFDT	338
QY	379	KEPAPTTPKEPAPNAPKKPAPPTPKPEAPTTPKPEAPTTTKE-----PSPTTPKPEAPTT	433
Db	339	ATETPFVAKSEDKMTLSKTAATETTOQTTEVTDGPEKETTKNVSIEIPITTVPLVEYTS	398
QY	434	TKSAP-----TTTKKEAPPT---TTKSAPPTPKEPSPTTKE---PAPT-----TP	472
Db	399	TSTASKESDGFHTLLKLVTTADSDSTESATTVKPFNEETTKSHVVPKPTKKGTVKTP	458
QY	473	K-----EPAPTTPKKPA-----TTTKE-----P	491
Db	459	KLELSFDEPTEIT-KAPHGCKLLEKKTYPHFVLSNDNFARYSEAKENDDYNHLDYNIYREAK	517
QY	492	APPTPKPEAP-----TTTTPKPAFTA--PKEPAPTT--PKETAPPTPKLTLPTTPEKLA--	540
Db	518	EPPTTESSTVEVTTTEEPANTGNPPTTENPTTTEQPTSTAESVTTALPTTTEQTVTTE	577
QY	541	PTTTPKAPPTPEELAPPTPEEPTTTPPEEPAPTTPKAAAPNTPKEPAPTTPKPEAPTT	599
Db	578	EPTTAEKSTATQ---KPTTQESVST---EKTSITKKA---SSTEE--PTTDEEPTTTT	625
QY	600	PKEPAPTTKETAPPTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSDKPAF-	658
Db	626	---ESSSTGTATTPELSTTSEETTTTELKITTE-----GSTTEEPTTTTAFEAEST	674
QY	659	---TTPKGAPTTTPKPEAPTTPKPEAPTTPKGTAPTTTLKEPAPTT-----	701
Db	675	GIITDDEETSTSTTTPETSITKE--IYTESAITQTSVVVESSTPRLPPEKWAIVNKF	732
QY	702	-----KKPAKELAPTTKGP-----TSTTSDKPAPTTPKETAPT	736
Db	733	KHNLEVLKCKKLLKEKSTSTGSDSETTTVAENIDEVTTTEKEKVQVTTITTEKS	792
QY	737	TPKEAPTTTPKPAFTTPB--TPBPTTSE-VSTPTTTEPTTIHKSDPESTPELSABPTP	793
Db	793	TTQOEETTTTTTTEKTSKTTTEKTEPTTSESATTEPTTSEPST-----TEST-----	838

Query Match 9.4%; Score 661; DB 2; Length 3507;  
Best Local Similarity 23.0%; Pred. No. 6.6e-21;  
Matches 278; Conservative 173; Mismatches 459; Indels 300; Gaps 53;

Query Match	9.4%;	Score 661;	DB 2;	Length 3507;
Best Local Similarity	23.0%;	Pred. No. 6.6e-21;		
Matches	278;	Conservative 173;	Mismatches 459;	Indels 300; Gaps
QY	27	LSSCAGCGGGYGRD-ATC-----NCDYNCQHYMECCPDFKRVCTAE-----	67	
Db	1781	MGSCGCKMAGYTGDCATCIKIEEPKSDKA-----CTDEWSRLCELEKKQCTVDEEV	1835	
QY	68	-----LSCGRCFSEFERGCDG-----DAQ-----	89	
Db	1836	PQCGACLPGHPHNGTCQSLQISGLCAQKNDCKNKAECIDIHPSHFCSCPDGFIGDGI	1895	
QY	90	CKKYDK-----CPDYESC-----AEVKDNKKVTKTKKTPKPPVVD	129	
Db	1896	CDDVDENNAGMCDDENTKCENTIGSFNCVCLGFKYDVCVYDEKKQPNREKIEDEE	1955	
QY	130	GSGLDN-GDFKVTT-----PDTSTTQHNKVSTSPKTTTA-----KPINRPSLPN	174	
Db	1956	NSSSNSNGQEKPTKGIIVSTASATSESTAEAPHVTTTSSITSTTKDMTGSKS-PENVTM	2014	
QY	175	SDTSKETSUTVNKETT-V-ETKETTNTTKOTSDGKEKTTSAKETQSTKTSAKDLAP---	230	
Db	2015	SSSEPVSTSSSKSTTASETTYSSTPSSSSSEAPLRTSSPATTTEVTESVSKSTTPKEE	2074	
QY	231	-TSKVIAKPTPKAETTTKGPALTTKPEPTPTTPKPEASTTTPKEPTTTIKS-APTTPKE	288	
Db	2075	SSSEITVKLLSSKSPVETSSVKSSPSTPS-ITTSQVSTVTPETSKSVLSSSEAPVSTSP	2133	
QY	289	APTITKSAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTTSKAPTTPKEPAPTTK	348	

Db 2134 TEVHT-SSEKPSLASSTGDNSTTPSTSSLASVKSTSAPEGTASVAPVKLSLSLSD 2192  
 QY 349 KPAPTTPK-BPAPTTPKEPTTPKPEAPTTKEP-APTTPKEAPAPKPKAPPTTPKEPA 406  
 Db 2193 VSQSTTKTFDATESSTVOAESTSGSVKSTSEPHVTKLSITSNPNSSVPTSPKST 2252  
 QY 407 PTTTKEAPATTTPKPSPTTP-KEPAPTTPKSAPTTPKPEAPTTKSAP-----TTPKE 458  
 Db 2253 PTVPE-----STEQPTSTFGQSLTPMNSNVEVLTSEPHVLSLSLSDVQSSTPN 2307  
 QY 459 PSPTTKEAPATTTPKPEAPTTKPKAPATTTPKPEAPTT-----503  
 Db 2308 LSESTVE-----TPKTSSEVSLNSEEPTTEAPTTPSDILSTTNLSSOSTVSTEDRS 2363  
 QY 504 -----TKKPAPTAPK-----EPAPTTPKETAPTTPKKLTPTTPKLAAPTTPKEP 547  
 Db 2364 EISSENSEKPT-SAPELVTSVTHVASSSDVPTESS-SEPDDL TGSSTENIPEASSKQT 2420  
 QY 548 APTTPEELAPTTPPEPTTPPEP-----APTTPKAAAPNTPKPEAPT-----591  
 Db 2421 ISSPTPTDTTASEPTTKSTMSDLSSTNSVLSESTTPESS-----KSPVSSSTEGIS 2476  
 QY 592 -----PKEAPT-----TPKEAP-----TTPKETAPTTPKGTAPTTPKPEAPT 630  
 Db 2477 VVTSTFESKVPSTLSSVLEEDLTTPSPILEETTTTASETSEPLETDSLTVSRIHELT 2536  
 QY 631 TPKPAPKELAPTTPKPTSTTSOKPA-----PT-----TP-K 662  
 Db 2537 TSENVPKESSTTSSESSKPSQEPAGILSTVWPTSSVSLITASEIEAITSNTPFKQ 2596  
 QY 663 GWAP-TTPK-----BPAPTTPKEAPTTPKGTAPTTPKPEAPTTPKPKAPKELAPT 713  
 Db 2597 GRTPITTPSKLVKSTTSPTSTVTSSEPTSESTKRTVSTVSTVSTTPTTEETTSLSLITAA 2656  
 QY 714 TKGPT-SITSOKPAPTTPKETAPTTPKPEAPTTPKPKAPTTTPPTTPPTTSVSTPTTKE 772  
 Db 2657 PSKPTSTSESEAPTTPAKTSKPSNVSTSRKSTENVTSTSQSLSSTMSSTS- 2715  
 QY 773 PTTIHKSPDESTP-ELSAEPTPKALE--NSPKPEGVPTTTPAKTK-----PEMTTAK 823  
 Db 2716 -----SEPETNAPAVTVSSEASSTLEENSTSSP-----TSSEASVLSLFPESITSEA 2766  
 QY 824 DKTERDLRTTPTTTPAAPTTPKATTTTEKTESKITATTT-----QVT 868  
 Db 2767 VTVSSR-----APAEITMSESHREISTVSSEPESEPEILSTVSPNVVTASSIPSEEPIL 2822  
 QY 869 STTODTTP-FKITT-----LKTTLAPKVTTTKTITTTTEIMNKPEETAKPKDRATNSK 922  
 Db 2823 SSVTSSSTPRVLITGTDPDLIVSVTVFSGHNRNONTASSV---PSNSTSPILPSES- 2878  
 QY 923 ATTPKQKTKAPKPTSTTKPKTPVRKPK-----TTPTRKMTSTMPEN-----PTSR 974  
 Db 2879 LTPQPPPTTTTAKPATTTSGRGPSPSIQPPAEMTTTAP-----PPPSNGGYGEETNQ 2932  
 QY 975 IAEAMLQTTT 984  
 Db 2933 EEEQVTSITT 2942

RESULT 12  
 T45025  
 mucin MUC5B, tracheobronchial [imported] - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T45025  
 R:Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.  
 J. Biol. Chem. 272, 3168-3178, 1997  
 A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat  
 A:Reference number: Z22899; MUID:97166151  
 A:Accession: T45025  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 1-3570 <DBS>  
 A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503  
 A:Experimental source: placenta  
 C:Genetics:  
 A:Gene: MUC5B

Query Match 9.2% Score 646.5; DB 2; Length 3570;  
 Best Local Similarity 22.5% Pred. No. 2.8e-20;  
 Matches 359; Conservative 107; Mismatches 536; Indels 591; Gaps 65;  
 QY 64 CTAE-----LSCGRCFESFERECDCDAQ-----CKYD---KCCPDYE 101  
 Db 506 CRAENYPEVSDVOGVLT-----SLETGLTCKNEQDTRFNCFNVRVLCDDY- 558  
 QY 102 SFCAEVKDNKKRTK-----KPTPKPVVDEAGSLDNGDKFVT--TPDTSITQ 149  
 Db 559 SHCPSTLATSTSTPSTPGTTWILTKPPTTATTTASTGSTATSTQNTACTPHVST- 617  
 QY 150 HNKVSTSPKITTA-----PINRPSLPNSDTSKETSITVNETTET-----KET 196  
 Db 618 ----ATPVTSSKATPFSSPGTATAPALRSTATTPTATSTATPSSSLGTTWRLSQT 673  
 QY 197 TTNKQTSDDGKEKTTSAKETOSIEKTS-----KDLAPTSKVLAKPTP- 240  
 Db 674 TTPMATMSTATSPSTPETHVSTVLTATTATTCATGSAVATPSSTPGTAHTTKVLTTTGG 733  
 QY 241 -----KAPT-----TTKGPAIT-----TPKEPTTTP-----K 263  
 Db 734 FTATPSSSGRAKTLPWISITTTTPTTRGSTVTPSSIPGTTHTPTVLTITTTTATGMSA 793  
 QY 264 EPASTTPKEPTTIKSAPTTPKEAPTTTKSAP--TTPKEAPTTT-KEAPT-----TP 316  
 Db 794 TPSSSTQTSPTPSLTTATTATGTTATGTTNPSTPGTTPPPVLTATTATTAATSTVTP 853  
 QY 317 KEAPTTTKEAPTTTKS-----APTTP-----KEAPTTPKP 350  
 Db 854 SSALGTHTPVPNTTATTGRLSPSSPHVCTATTSATSGILGTHTEPSTGTHSHPT 913  
 QY 351 APTT----PREAPTTTPKEPTTPTPKEAPTTTPKEAPTTAPKPKAPTTTPKEPA 406  
 Db 914 AATGTQHTSPALSSPHSPSRSTTESPPSGTTTPGHTT-----ATSRRTATATPSKR 967  
 QY 407 PTT--PKEP--APTTHK-----EPS-----422  
 Db 968 TSTLLPSQTSAPITTVVTMGCEPQCAWSEWLDYSYMPMPGSGDDFTTSNIRAGAVNC 1027  
 QY 423 -----422  
 Db 1028 EQLGLECRAQAQGVPLRELQGVVECSLDFGLVCRNREQVKFKMCFNVEIRVFCNYG 1087  
 QY 423 --PTTPKEAPTTTKSAPTTT--KEAPTTTKSAPTTTPKEPSTTPKEAPTTTPKEAPT 478  
 Db 1088 HCPSTPATSSATPSPTPGTTWILTELTATTTESTGSTATPTSLRTAP--PPKVLTT 1145  
 QY 479 TPKPAPTTPKEAPTTTPKEAPTTTKKA-----TTPAK-EPAP-----517  
 Db 1146 TATTPTVTSSK-----ATPSSPGTATAPALRSTATTPTATSVTPIPSSSLGTTWRLSQ 1201  
 QY 518 -TTPKETAPTTPKKLTPTTPE-----KLAPTPEKAPTTPEELAPTTPPEEPTTPTEE 570  
 Db 1202 TTTPTATMSTATSPSTPETAHTSTVLATATTGATGSAVATPSSTPGTAHTTKVPTTTT 1261  
 QY 571 PAPTTPKAAAPTTPKEP-----APTTPKEAPT--TPKEAPTTTPKETAPTTPKTAAPT-T 623  
 Db 1262 GFTATP--SSSGTALTTPPVWISITTTTPTTRGSTVTPSSIPGTTHTATVLTITTTTATGS 1320  
 QY 624 LKEAPTTTPKPKAPKELAPT-----TTKEPTSTTSKAP-----TTPKGTAPT- 667  
 Db 1321 MATSSSTQTSPTPSLTTATTATGTTNPSTPGTDRIPPVLTATTATTAATSTV 1380  
 QY 668 TPKEAPTTTPKEAPT-----683

```
Db 1381 TPSSALGTHTPVPNTTATTHGRSLSPSSPHTVRTAWTSATSGLTGLTHITPSTGTSH 1440
Qy 684 TPKGTAFTT-----LKEAP---TPPKPAKELAPTTTKGTSSTSDKPAFTTKETA 734
Db 1441 TPAATGTTQHSTPALSSPHSRTTESPPS---GTTTGHITATSRATATPSKTR 1496
Qy 735 PTT--PKP-----741
Db 1497 TSTLLPSSPTSAPITTVVTWMCBPCAWSEWLDYSYMPGPGSGDDFTYSNIRAAGAVC 1556
Qy 742 -----741
Db 1557 EQPLGLECRAQAGQVPLRELQGVBCSLDGLVCRNREQVGFKMCFNVEIRVFCNYG 1616
Qy 742 -APTTKKKAPPTPEPPPTT---SVSTPTTTKEPTTHHSDESTPELSAEPKALE 797
Db 1617 HCPSTATSTPSTPGTWTLLTQTTAAATTAATGTAIP--STP--GTAPPKVL 1673
Qy 798 NSPKPGVPTTKPAATKPEMTT-----AKDKTTERDLRTTPTT-----838
Db 1674 SOATTPATSSKATSSSPRTATTLPLVLTSTATKSTATSFTPIPSSTLGTGTSQNRPPH 1733
Qy 839 -----TAAPKMKETA-TTETKTESKITATTTQVTSI--TTQDTP-----FKITLTK 884
Db 1734 PMATMTIHPSSPTTHTSTVLTK-----ATTTRATSSMSTPSTPGTTWILTELTAA 1789
Qy 885 TTTLAPKVTTKKTIITTEIMNKPEETA-----KPKDRAT--NSKATTPKQPK 931
Db 1790 TTTAALPHGTPSTPGTWTLLTPESTATVTVTGSTATASSTRATAGTLKVLSTATTTP 1849
Qy 932 TKAPKPTSTKKPKT---MPVRKPKTTPPKMTSTMPELNPTSRIAEA--MLQTTTRP 986
Db 1850 TVISSRATPSSSPTATALPALRSTATTPTATSVTAI-----PSSSLGTAWTRLSTQTTT 1904
Qy 987 NQTPNSKLVENPKSDEGAGAEETHMLLRPHVHPEVTPMDYLPVP---NOGIILN 1043
Db 1905 TATMSTATPSTPETVHTSTVLTTATTTGTGSVATPSSPTPGTAHTKVPSTTTGTAT 1964
Qy 1044 PMLSDSTNLCNGKPPVDGLTLRLNGTLVAPRGHYFWMLSPPSPSPARRI-----1092
Db 1965 PSSSPTALT--PPVWIST---TTPTRGS---TVTSSIPGTHTATVLTITTTTVA 2015
Qy 1093 -----TEVWGIPIPIDVTRCNCBGT 1115
Db 2016 TGSMAFPSSSTQTSSTGTPSLTTTATTITATGST 2048

RESULT 13
I38346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38346
R:Labell, S.; Koilmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330
A:Accession: I38346
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 9.0%; Score 635.5; DB 2; Length 7962;
Best Local Similarity 22.8%; Pred. No. 1.8e-19;
Matches 306; Conservative 112; Mismatches 475; Indels 451; Gaps 63;
Qy 76 ESPERG-----RBCDCAQCKKYDKCCPDYE--SFCAEVKDNKKNR-----TKKKTP 121
```

```
Db 5783 EDFEEQYVEREGYDEGEEMEEAYQREVIOVQKEVVEESHKRVPAKVPEKKAPPP 5842
Qy 122 ----KPPV---VDEAGSLDNGDFKVT-----TPDTSSTQHKNKVSYS-PKITT 161
Db 5843 PKVIKRPVIEIKETISRRMEEEKVQVTKVPEVSKKIVPOKPSRTPVQEEVIEVKVPAVHT 5902
Qy 162 AKPINRPSLPPNSDTSKETSLSL---VNETTVETKETTTTNKOTSDGKEKTTSAKETQ 218
Db 5903 KKMVISEEKMFFASHTEEEVSVTPVQKEIVTEEKIHHAVSKRVE-----PPPKVPE 5955
Qy 219 SIEKTSAKDLAPT--SKVLAKPTPKAETTTKGPALTTTPKEPTTTPKPEAS-----TTPK 271
Db 5956 LPEKPAPEAVAPVIPKVPPEPAKVPVEPKP--VPEEKKVPVPKPEAPAAKVPKVPVK 6014
Qy 272 EPTPTTIKSAPTTPKEPAP-----TTTKSAPTTPKEPAPTTTKEPA--PTTPKE 318
Db 6015 KVPPEEKIPVPVAKKEAPPAKVPQGVVTEEKITIVTQREESPP---PPAVPEIPKK 6070
Qy 319 PAPTTPKPEAPTTTKSAPTTPKPEAPTTPKKPAFTTP--KEPAPTTPKEPTT-----370
Db 6071 KVPPEERK--VPKREEEVPPPKVPA--LPKKVPPEEKVAVVPVAKKAPPPRAEVSKTV 6127
Qy 371 -----PKEPAPT 377
Db 6128 VEEKRFVAEEKLSFAPVQREVTRHEVSAEESYSYSEEEGVSVISVYREEREEREEAEV 6187
Qy 378 TK-----EPAPT-----PKPA-----PT 392
Db 6188 TEYVAMEPEEYVVEEKLHIISKVPAEPAEAVTERQEKIIVLKPPIAKIIEEPPPAKVP 6247
Qy 393 APKK-----PAPTTPKE--PAPTTPKEP--APTTPKEPSPTTPKE---PAPTTPKSAP 438
Db 6248 APKKIYPEKKVPAVPKKEKVPVPKVPPEEKKVPPEKVPKVPKVIKMEEPILAKVTEKHM 6307
Qy 439 TTTKEP--APTTPKSAP---TTPKEPSPTTKE-----PAPT-----471
Db 6308 ITQEEKVLVAVTKKAPPAKVPPEEKRAVPPEEKVLKPKREEEPPAKVTEPRKRVKE 6367
Qy 472 -----PK--EPAP-----477
Db 6368 EKVSIEAPKREPQPIKEVTIMEEKERAYTLLEEAVSVQREEEYEEYEDYKEFEYEPT 6427
Qy 478 -----TTPKKKAPT--TPKEPAPTTPKPEAPTTTKKKAPTA 511
Db 6428 EYDQYEEYEEYEEYEEYITEPEKIPVKEPVEEPVPVTPKPAKPAKVLKAVP-E 6486
Qy 512 PKEPAPTTPKETAPTTPKLTPTTPEK-----LAPTTPKEPAPTTPPEELAPTPEEPT 564
Db 6487 EKVPVP--IPKKLKPPPPK--VPEEPKVPPEEKIHSITKREKEQVTP---AAKVPMPK 6540
Qy 565 PTPPEEPAPTTPKAAAP--NTPKEPAPTTPKEPA-----596
Db 6541 RVVAEEKVPVPKREVAPPVVRVPEVPKLEPEEVAFEVEEVTVHVEEYLVVEEEYIHEEEE 6600
Qy 597 PTPPKPEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPAKPAKELAPTTTKEPTSTT---652
Db 6601 FITEEEVVPVIPVK--VPEVPKVPPEE--KRPVPVPKKEAPPKVPVDPVKKPEEKVPLI 6658
Qy 653 --SDKPAPTTPKGTAPTTPKEPAPTTPKGTAPTTPKGTAPTTLKEPA--PTTPKKKAPKE 708
Db 6659 PKKEKP-----PPAKVPPEVPKVPV--EEKVPVPVPPKKVEAP-----PAKVPPEVPKVP 6708
Qy 709 L-----APTTPKGTSTTSOKPAPTTPKETAPT--TPKE---PAPTTPKKKAPAPTTPETPPT 760
Db 6709 KVPVPAPKKVVEAPPKVPPEVKLLIPEEKKPTVPKKEAPPKVPKREPPVPVPVVALPQ 6768
Qy 761 TSEVSIPTTTTKEPTTIHKSPEDESTPELSAEPKALENSKEPCVPTTKTKPAATKPEMTT 820
Db 6769 EEEVLFEETIVPEEVLPEEEVLPPE--EEVVLPEEEVLPPEEEIPEEEVEEVPPEEYVP 6827
Qy 821 TAKDKTTERDL--RTTPTTTTAAPKMTKETATTTETKTESKITATTTQVSTTTTQDTPF 878
```

Db 6828 EEEFVPEEVLPEVKVPVPAP-----VPEIKKKVTEKKVVIPKKEEAAPAKVPEVPK 6882  
QY 879 KIITLTKITLAPKVTTTKKTIITTEIMNKPEETAQPKDRATNSKATTPKQKTKAPKPK 938  
Db 6883 KVE--EKRIILPK-----EEEVLPEVTEEPPE-----EPISSEETPEEP 6920  
QY 939 TSTKK-----PKTMPVRVK---PKT-TPTPRKMTSTMPELNPTSRIAEMLQITTRNQ 989  
Db 6921 PSTEEVEAVPRVPEVIKKAPEAPTPVKVEA--PPAKVSKKIPPEKVPVPVQKKEA 6978  
QY 990 PNSKLVENVKSEDAGAGETPHMLRPHVFMPEVTPDMYILPRVNOGIINPMLSDE 1049  
Db 6979 PPAKVPEV-----PKKVPKVKVL--VPKKEAV-----7003  
QY 1050 TNICNGKPDGLTLRNGTLVAER 1073  
Db 7004 -----PPAKGRTVLEEKVSVAFR 7021

RESULT 14  
T11622  
extensin class 1 precursor - cowpea  
C:Species: Vigna unguiculata (cowpea)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 17-Nov-2000  
C:Accession: T11622; S54155  
R:Arsemajevic-Maksimovic, I.; Broughton, W.J.; Krause, A.  
Mol. Plant Microbe Interact. 10, 95-101, 1997  
A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.  
A:Reference number: Z17301; MUID:97155574  
A:Accession: T11622  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-489 <RS>  
A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937  
A:Experimental source: sub-species Red calcoona  
R:Arsemajevic-Maksimovic, I.; Broughton, W.J.; Krause, A.  
submitted to the EMBL Data Library, April 1995  
A:Description: A class of root-hair specific extensins involved in rhizobium/legume inte  
A:Reference number: S54155  
A:Accession: S54155  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 326-489 <AR2>  
A:Cross-references: EMBL:X86030; NID:g791149; PID:g791150  
C:Gene: Ext263  
C:Superfamily: hydroxyproline-rich glycoprotein  
C:Keywords: glycoprotein; hydroxyproline  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-489/Product: extensin class 1 #status predicted <MAT>

Query Match 9.0%; Score 633; DB 2; Length 489;  
Best Local Similarity 31.2%; Pred. No. 1.6e-20;  
Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 7;  
QY 285 PKEPAPTTTKSAPTTPKEPAPTTTKPAPTTTKPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAP 344  
Db 39 PKOTPPYYNAPYYKSPSP-----PSPSP-----PPVYVHKYPPYYKSPSP 83  
QY 345 TTPKKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 401  
Db 84 PSPSPPPYYKSP 141  
QY 402 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 461  
Db 142 YKSP 137  
QY 462 TTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 521  
Db 188 YYKSP 247  
QY 522 ETAPTTPKLTTPTEKLAAPTTPEKAPTTPEELAPTTPEELAPTTPEELAPTTPEELAPTTPE 581

Db 248 PPPYYKSP 307  
QY 582 NTPEKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 641  
Db 308 PSPSPPPYYKSP 366  
QY 642 PTTTKEPSTSTSDKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 701  
Db 367 ----KSPPSP 422  
QY 702 KKPAPKELAPTTTKGTSTTSDKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 759  
Db 423 SPP-----PPVYKSP 475  
RESULT 15  
C84672  
hypothetical protein At2g27380 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: C84672  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: C84672  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-761 <STO>  
A:Cross-references: GB:AE002093; NID:g5306260; PIDN:AAD41992.1; GSPDB:GN00139  
C:Gene: At2g27380  
A:Map position: 2

Query Match 8.9%; Score 632; DB 2; Length 761;  
Best Local Similarity 31.1%; Pred. No. 2.7e-20;  
Matches 217; Conservative 44; Mismatches 354; Indels 82; Gaps 33;  
QY 238 PTPKAETTTKGPALTTTP---KEPTPT-TPKEPASTTPKEPTPTTKSAPTTPKEPAPTTT 293  
Db 69 PPPIQKPPTYSPIYPPPIQKPPTYSPIYPPPIQKPPTYSPIYPPPIQKPPTYSPIYPPPIQKPPT 128  
QY 294 KSAPTTTPKEPAPTTTPKEPAP--TTPKEPAPTTTPKEPAPTTTKSAPTTP---KEPAPT-- 345  
Db 129 YSPPIYP-----PPPIQKPPTYSPIYPPPIQKPPTYSPIYPPPIQKPPTYSPIYPPPIQKPPT 181  
QY 346 TPKKP---APTTPKEPAPTTTP---KEPTP--TTPKEPAPTTTPKEPAPTTTPKEPAPTPAK 395  
Db 182 PPVKPPVHKPPTPIYSPPIKPPVHKPPTPIYSPPIKPPVHKPPTPIYSPPIKPPVHKPPTPIYSP 241  
QY 396 KPAPT-TPKEPAPTTTPKEPAP---TTTKEPSTTPTPKEPAPTTTKSAPTTPKEPAPTTTKS 451  
Db 242 PPTPIYSPPIKPPVHKPPTPIYSPVVKPPVQTPPTPIYSPVVKPPVHKPPTPIYSPV 301  
QY 452 APPTKEPSTTTTKEP---APTTPKEPAPT--TPKKAPAPTTTPKEPAPTTTPKEPAPTTTK 506  
Db 302 VKSPVQKPPTPIYSPPIKPPVQKPPTPIYSPPIKPPVQKPPTPIYSPPIKPPVQKPPTPIYSP 360  
QY 507 PAF--TAPKEPAPT---TPKETAPTTP---KKLTPTTPEKLAAPTTTPKEPAPTT---PE 553  
Db 361 PPIYSPVVKPPVHKPPTPIYSPVVKPPPIQKPPTPIYSPVVKPPPIQKPPTPIYSPVVKPP 420  
QY 554 ELAPTTPEEPTTPTEEPAPT---TPKAAAP-NTPEKAPAPTTTPKEPAPTTTPKEPAPTTTP 608  
Db 421 KLPPVKKPPTPIYSPVVKPPVHKPPTPIYSPVVKPPVHKPPTPIYSPVVKPPVHKPPTPIYSP 480  
QY 609 KETAPTTTPKGTAPTTTPKEPAPT--TPKKAPKELAPTTTPKEPTSTTSKDPAPTTPKGTAP 666  
Db 481 TYSPPVQV-----PPVOKPPTPIYSPVVKPPPIQKPPT---PTYSPPPIKPPVVKPPPTPT 532











```
QY 529 -----KKLTP----- 533
Db 1789 TGWLDGKPNFHKPGGDTLIGVCGPGWAANISCRATMYDPVPIQGLQGTVVCDVSVGL 1848
QY 534 -----TPKELAPTTPEKAPATTPEELAPTTPEE 562
Db 1849 ICKNEQKGGVPMFACFLNYEINQCCBVTQPTMTTITENTENPTTPTTPTTPTT 1908
QY 563 PPTT-TPPEAPATTTPKAAANTPKBPATTPEK-----APTTPKEPAPT-TPKEAPT 615
Db 1909 PPTTPTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1968
QY 616 PKGTAPTTLKEPAPTTPKKAPKAPKELAPTTTKEPTSTTSKDPAP-TTPKGTAPTTPEKAP 674
Db 1969 PTTTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2028
QY 675 TTPKEPAPTTKGTAPTTLKEPAPTTPKKAPKAPKELAPTTTKEPTSTTSKDPAP-TTPKET 733
Db 2029 TGTQTP-TTPTTTTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2087
QY 734 APTTPKEPAPTTPKKAPATTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 793
Db 2088 TTVTPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2143
QY 794 KALENSPKP-----GVPT-TKTPAATKPEMTTAKDKTTERDLR---TTP-ET 837
Db 2144 TGTQTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2203
QY 838 TTAAPKMT---KETATTEKTTESKITATTTQVSTTTQDTPPTPKITLTKTLAPKVT 894
Db 2204 TTVTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2258
QY 895 T-KTITTEIMNKPEETAKPKDRATSKATTPKQPKAP-KKPTSKKPKTMDPRVK 952
Db 2259 TGTQTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2318
QY 953 -PKTTPKMTSTWPELNP---TSRIEAMLOTTR-PNQTSPSKLVEVNPKSEDAGGAE 1008
Db 2319 TTVTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2375
QY 1009 GETPHMLLRPHVFEVPT 1027
Db 2376 TQTP-TTPTTTTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2375

RESULT 2
SLPI_CLOTH
AC Q06852; STANDARD; PRT; 1664 AA.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER PROTEIN 1).
DE OLPB.
GN Clostridium thermocellum.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=9320931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein CipA and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
CC - SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC - SUBCELLULAR LOCATION: CELL WALL.
CC - SIMILARITY: CONTAINS 4 S-LAYER HOMOMOLOGY (SLH) DOMAINS.
CC -----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; X67506; CAA47841.1; -
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW Cell wall; S-layer; Signal; Repeat.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
FT DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.
FT REPEAT 36 191 1.
FT REPEAT 207 383 2.
FT REPEAT 409 585 3.
FT REPEAT 607 763 4.
FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
FT T-P-S-D-E-P.
FT DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.
FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).
FT DOMAIN 1495 1565 SLH 2.
FT DOMAIN 1566 1625 SLH 3.
FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).
FT SEQUENCE 1664 AA; 178194 MW; 5F396695BA9F74B CRC64;
```

```
Query Match 13.4%; Score 950; DB 1; Length 1664;
Best Local Similarity 31.0%; Pred. No. 3.5e-32;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;
```

```
QY 234 VLAKTP-KAETTTKGPALTTPKEPTTPTPKEPASTTPKEPTTPKAPTTKAPTTKEPAPT 292
Db 758 VVIQAPAKAASDEIPTDTPSDEPTPS-----DEPTPS---DEPTPSDEPTPSD 804
QY 293 TKSAPTTPKEPAPTTPKEPAPTTPKEPA-----PTTTPKEPAPTTPKAPTTKAPTTK 348
Db 805 EPTPSETPEEPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPTDTP 864
QY 349 KPAPTTPKEPAPTTPKEPTTPTPKEPAPTTPKEPAPT-TPKEPAPTAPKAPTTKKEPAP 407
Db 865 SDEPTPSDEPTPS--DEPTPS--DEPTP-SDEPTSETPEEPTDTPSDEPTPSDEPTP 919
QY 408 TTPKEPA--PTTTPKEPST-TPKEPAPTTPKS-APTTPKKEPAPT---TTKSAPTTPKEPS 460
Db 920 SDEPTPSDEPTPSDEPTPSDEPTPEEPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 979
QY 461 ---PTTTPKEPAPT-TPKEPAPTTPKAPTTTPKEPAPTTPKEPAPTTPKAPTTKAPTTAPKEP 515
Db 980 PSDEPTPSDEPTSETPEEPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEP 1039
QY 516 APT---TPKETAPTTPKALTTPTPKEPAPTTPKEPAPT---TPEELAPTTPEEPTPT-TP 568
Db 1040 TSPDEPTPSE---TPEEPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSD-EPTPSDEPTSETP 1094
QY 569 EEPAPTTPKAAANTPKBPATTTPKEPA--PTTTPKEPAPT-TPKETAPTTPKGTAPTTLK 625
Db 1095 EEPITDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPTDTPSDEPTPSD 1154
QY 626 EPAPTTPKAPKELAPTTTKEPT-STTSKAPATTTPKGTAPTTTPKEPAPTTPKAPTTPEKAP 684
Db 1155 EPTPS--DEPTPSD-EPTPSDEPTPEEPTDTPSDEPTPSDEPTPS--DEPTPS- 1208
QY 685 PKGTAPTTLKEPAPT-TPKAPKAPKELAPTTTKEPTSTTSKAPATTTPKETAPTTPKE 740
Db 1209 ---DEPTPSDEPTPEEPTPEEPI-----PTDTPSDEPTPSDEPTPSD-EPTPSDE 1253
QY 741 PAPTTPKAPAPT-TPETPTPTTSEVSTPTTTPKPTTHKSPDESTPELSAETTPKALENS 799
Db 1254 PTPS--DEPTPSDEPTPEEPTDTPSDEPTPSDEPT-----PSDEPTP--SDEPTPSDEPTP 1305
```





[12]
SEQUENCE OF 1-46 FROM N.A.
TISSUE-Breast carcinoma;
Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CYTOSKELETON.
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
IS ALSO PRODUCED.
-1- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
SPLICING.
-1- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
TUMORS, SUCH AS BREAST CANCER.
-1- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND STALIC
ACID).
-1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT
VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
-1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).
EMBL; J05582; AAA60019.1; -
EMBL; M32738; AAA35804.1; -
EMBL; M32739; AAA35806.1; -
EMBL; J05581; AAA59876.1; -
EMBL; M61170; AAB53150.1; -
EMBL; X52229; CRA36478.1; ALT\_SEQ.
EMBL; X52228; CRA36477.1; ALT\_SEQ.
EMBL; M35093; AAB59612.1; ALT\_SEQ.
EMBL; Z17324; CAA78972.1; -
EMBL; Z17325; CAA78973.1; -
EMBL; M31823; AAA35757.1; -
EMBL; S81781; AAD14376.1; ALT\_INIT.
EMBL; S81736; AAD14369.1; ALT\_INIT.
EMBL; M21868; AAA59874.1; ALT\_SEQ.
PIR; A35175; A35175.
PIR; B35175; B35175.
PIR; S10218; S10218.
GlycoSuiteDB; P15941; -
MIM; 158340; -
MIM; 113720; -
InterPro; IPR000082; SEA.
SMART; PF01390; SEA; 1.
SMART; SM00200; SEA; 1.
PROSITE; PS50024; SEA; 1.
Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
Repeat; Alternative splicing.
FT SIGNAL 1 23
FT CHAIN 24 1255
FT DOMAIN 24 1162 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1163 1186 POTENTIAL.
FT DOMAIN 1187 1255 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 81 960 44 X 20 AA TANDEM REPEATS.
FT DOMAIN 1034 1151 SEA.
FT CARBOHYD 957 957 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 975 975 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1133 1133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 19 19 T -> TAITAPKAT (IN ISOFORM B).
FT VARSPPLIC 20 22 MISSING (IN ISOFORM C).
FT VARSPPLIC 20 31 MISSING (IN ISOFORM D).
FT VARSPPLIC 126 905 MISSING (IN POLYMORPHIC EPITHELIAL
ISOFORM).
FT VARSPPLIC 1077 1087 FLQIYKGGFL -> VSIGLSPMLP (IN SECRETED
ISOFORM).

FT	VARSPPLIC	1088	1255	MISSING (IN SECRETED ISOFORM).
FT	CONFLICT	2	2	T -> A (IN REF. 11).
FT	CONFLICT	134	134	P -> Q (IN REF. 9).
FT	CONFLICT	154	154	P -> Q (IN REF. 9).
FT	CONFLICT	1021	1021	S -> T (IN REF. 3).
FT	CONFLICT	1251	1251	A -> T (IN REF. 3).
SQ	SEQUENCE	1255	AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;	

  

Query Match	9.2%;	Score 651;	DB 1;	Length 1255;
Best Local Similarity	27.8%;	Pred. No. 3.9e-20;		
Matches 295;	Conservative 72;	Mismatches 460;	Indels 234;	Gaps 51;

  

QY	181	TSISVTNKKETTETTTNKOTSTDKGKTTSAKETOSIEKTSAKD-LAPTISKVLAKPT	239
DB	16	TVLTV-----VTGSGHASTPGEKETSATQSSVPSSTKNAVSMTSSVLSSH	65
QY	240	P-KAETTTKGP-ALTTTPKEP-----TPPTPKPEASTTKPEPTTII---K	279
DB	66	PGSGSTTQGDVTLAPATEPAGSAAATWGQDVTSVPVTRPALGSTTPPAHDVTAPDNK	125
QY	280	SAPTTKPEAPITTKSAPTTKPEAPITTKPEP-----APTTPKPEAPITTKPEAPITTK	333
DB	126	PAPGSTAPPAHGVTSAPDT--RPAGGSTAPPAHGVTSAPDT--RPAGGSTAPPAHGVTS	179
QY	334	SAPTTKPEAPITTP-----KKPAPTTKPEP-----APTTPKPEPTTP-----	371
DB	180	SAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS	239
QY	372	-----KEPAPTTKPEP-----APTTPKPEAPITAPK-----KPAPTTKPEAPITTKP	412
DB	240	SAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS	299
QY	413	PAPTTKPEAPITTKPEAPITTKSAPTTKPEAPITTKSAPTTKPEAPITTKPEPPTTKP	467
DB	300	SAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS	356
QY	468	-----APTTPKPEAPITTKP-----KPAPTTKPEAPITTKP-----EPAPTTTKKAPTA	515
DB	357	GVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAH	416
QY	516	APTTPKPEAPITTKPPTTPEKLAPTTPEKLAPTTPEKLAPTTPEEPTTPEEPTTPEE	575
DB	417	GVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAP	472
QY	576	PKA-----AAPNTKPEAPITTKP-----EPAPTTKPEAPITTKPETAITTKGTA	627
DB	473	PPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAG	528
QY	628	APTTP-----KKPAPKELAPITTKPTTSTSDKPAP---TTPKGTAPITTKPEAPT	675
DB	529	GSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR	583
QY	676	TPKEPAPITTKP-----TAPITTKPEAPITTKP-----KKPAPKELAPITTKG	721
DB	584	TRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAP	642
QY	722	SDKPAP---TTPK-----ETAPITTKPEAPITTKP-----KKPAPITTKPTT	765
DB	643	DTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS	700
QY	766	TPITTKPEPT-----IHKSPDESTPELSAETPKALENSPKPEGVPTTKTTPA	818
DB	701	APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAH	753
QY	819	TTAKDKTTERDLRTTPPTT-----TAAPKMKETATTTTEKTESKITATTTQV	871
DB	754	--PAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR	807
QY	872	IQDITTP--FKITITLTKTTTLAKPVTTTKKTIITTT--EIMNKPEETAKPKDR	927
DB	808	PGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR	867





```

QY 735 PTPKKEPAPT 744
Db 967 PASCKSPRS 976

RESULT 7
EXTN_TOBAC
AC P13983;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
GN HRGPN3.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. XANTHI; TISSUE=Leaf;
RX MEDLINE=90128263; PubMed=2612909;
RA Keller B., Lamb C.J.;
RT "Specific expression of a novel cell wall hydroxyproline-rich
glycoprotein gene in lateral root initiation.";
RL Genes Dev. 3:1639-1646(1989).
CC -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
MAIN ROOT.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
GLYCOSYLATED.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: X13885; CAA32090.1; -
P.R.: S06733; S06733.
Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
Hydroxylation.
SIGNAL 1 ?
CHAIN ? 620 EXTENSIN.
REPEAT 70 73 H-A-P-P.
REPEAT 148 151 H-A-P-P.
DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
REPEAT 229 235 1.
REPEAT 236 242 2.
DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
SEQUENCE 620 AA; 65406 MW; 641DD278AB28524 CRC64;

Query Match
Best Local Similarity 7.8%; Score 551; DB 1; Length 620;
Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

QY 242 AETTK-GPALTTP--KETPTT-----PKPASTTPKE---PTPTTKSAPTTPKEPAP 290
Db 24 AEATQGYGLPPVTSQPPSSIGLSPSAPTTPPSRGHVPSP---RHAPRHAYPPP 80
QY 291 TTTKSAPTTPKEP-----APTTPKEAPT---TPKEAPTTPKEAPTTPKSAPTTPKE 341
Db 81 SHGHLPPSVGGPPHGRHLPGRGNGNPPSPVSPSHPPSYGAPPSPSHGHLPSHGQR 140
QY 342 PAPTTPKKPAPT---TPKEPAPTTPKEPTTPTPKKEPAPT---EPAPTTPKEAPTAPKK 396

```

```

Db 141 P-----PSPSHGAPPSCGHTPPRGQHPSPSHRRPSPSPSRHGHPPTTYAQPPTPIYS 193
QY 397 PAPTTPKEAPTTPKKEPAPTTPKESPTTPK---EPAPTTPKSAPTTPKKEPAPTTPKS-- 451
Db 194 PSQV--QPPPT--YSPPPPTHVQPTSPPSRGHQPOPTTHRHAPPTHRHAPTHQDSPL 249
QY 452 ---APTTPKESPTTPKKEPAPTTPKKEPAPTTPKKEPAPT---TPKKEPAPTTPK 505
Db 250 RHLPPSPRRQPPPTYSPPPPAYAQSPQSPPTYSPPPTYSPPPPSYSPPPAYSPSP 309
QY 506 KPAPT---APKEPAPTTPKETAPTTPKKL-TPTT-----PEKLAPTTPKKEPAPTPEEL 555
Db 310 PPTPTPTSPPPAYSPPTYSPPPTTYLPLSPSYSPPPSYSPPPSYSPPPSYSPPTYL 369
QY 556 APITPEEPTTTEEPAPTTPKAAAP---NTPKEPAPTTPKKEPAPTTPKKEPAPTTPKETA 612
Db 370 PPPPPSPPPSPSPPTTVEQSPPPPPAYSPPLPAPTYSPPPT--YSPPPPTVAQPP 427
QY 613 PTTPKGTAPTTLKKEPAPTTPKPKAPKELAPTTPKPTSTTSDKAPAPTTPKGTAPTTPKEP 672
Db 428 P-----LPPTYSPPPPAYSPPTYSPPPTYSPPPPAYAQPPPPPTYSPPPPAYSP 481
QY 673 APTTPKEPAPTTPKGTAPTTLKKEPAPTTPKPKAPKELAPTTPKGTSTTSDKAPAPTTPKE 732
Db 482 PPPSYSPPPPPQVQLPPTTFSPPPRRIRHLPPPHRQP---RPPTTYGQPPSPPTFSP 538
QY 733 TAPTTPKEAPT---TPKKAPAPTTPPTTSEVSTPTTKEPTTIHKSPDSTPESLAE 790
Db 539 PPRQIHSPPPHWPQRTPTTYGQPPSPPTTSAPPRIHSPPPHRRQPPRPTTYGQP 598
QY 791 PTPKALENSPKPGVPTTKTPAATKP 816
Db 599 PSP-----PTTYSPPSPPP 612

RESULT 8
NFH_MOUSE
ID NFH_MOUSE STANDARD; PRT: 1087 AA.
AC P1246; Q61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN NEFH OR NEFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89121513; PubMed=3220257;
RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
RA Mushynski W.;
RT "Sequence and structure of the mouse gene coding for the largest
neurofilament subunit.";
RL Gene 68:307-314(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89089138; PubMed=3145094;
RA Sheldman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
RT "The structure of the largest murine neurofilament protein (NF-H) as
revealed by cDNA and genomic sequences.";
RL Brain Res. 464:217-231(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
RA Carden M.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT

```





RX MEDLINE-91376547; PubMed-1896773;  
RA Prioli R.P., Mejia J.S., Aji T., Akawa M., Pereira M.E.A.;  
RT "Trypanosoma cruzi: Localization of neuraminidase on the surface of  
RL Trypomastigotes";  
CC Trop. Med. Parasitol. 42:146-150(1991).  
CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN  
CC PARASITIC INVASION OF CELLS.  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC  
CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLEURAMINYL  
CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLEURAMINYL  
CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,  
CC GLYCOLIPIDS OR COLOMINIC ACID.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
CC (POSSIBLE).  
CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM  
CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.  
CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT  
CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA  
CC PROTEIN.  
CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.  
CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M61732; AAA30255.1; -  
CC PIR: JH0557; JH0557.  
CC HSSP: P29768; 101L.  
CC InterPro: IPR002860; BNR.  
CC Pfam: PF02012; BNR; 2.  
CC Hydrolase: Glycosidase; Glycoprotein; Repeat; GPI-anchor;  
CC phosphorylation.  
CC DOMAIN 1 457 CYS-RICH.  
CC REPEAT 23 34 BNR 1.  
CC REPEAT 163 174 BNR 2.  
CC REPEAT 209 220 BNR 3.  
CC DOMAIN 458 588 FIBRONECTIN TYPE-III.  
CC DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.  
CC CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SEQUENCE 1162 AA; 07049221897C6A40 CRC64;  
CC  
CC Query Match 7.5%; Score 530.5; DB 1; Length 1162;  
CC Best Local Similarity 29.4%; Pred. No. 2.9e-15;  
CC Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;  
CC  
CC QY 258 TPTPKPKE-PASTTTPKEPTTTSAPTTPKEAPTTPKSAPTTPKAPATTPKAPATTPK 316  
CC Db TPSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS 551  
CC QY 317 KE-PAPTTPKAPTTPKSAPTTPKE-PAPTTPKPP-----APTTPKEA-----PTTP 363  
CC Db ADSAAGTSTPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS 711  
CC QY 364 KEPTT-TTPKEAPTTPKAPTTPKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPK 422  
CC Db VDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSA 764  
CC QY 423 PTPPKAPTTPKSAPTTPKAPTTPKSAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 478  
CC Db HGTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSP 817  
CC QY 479 TP-KKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 533  
CC Db TPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSA 876

QY 534 TTP-EKLAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 583  
Db STPVDS--SAHSTSTPADSSAHS--TPSTPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHST 935  
QY 584 PKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 643  
Db PSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSP 983  
QY 644 TTKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 702  
Db PSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSP 1033  
QY 703 KPAPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 758  
Db TP-----ADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHS 1084  
QY 759 PTTSEVSTPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 817  
Db DSSAH-STPTPAD--SAHSTSTPADSSAHSPTSPVDS--SAHSTSTPADSSAHSPTSPVDS--SA 1125  
QY 818 MT 819  
Db 1126 GT 1127  
  
CC RESULT 10  
CC CPN\_DROME STANDARD; PRT; 865 AA.  
CC AC Q02910;  
CC DT 01-OCT-1993 (Rel. 27, Created)  
CC DT 01-OCT-1993 (Rel. 27, Last sequence update)  
CC DT 01-FEB-1994 (Rel. 28, Last annotation update)  
CC DE CALPHOTIN.  
CC GN CPN OR CAP.  
CC OS Drosophila melanogaster (Fruit fly).  
CC OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
CC OC Pterygota; Neoptera; Endopterygota; Diptera; brachycera; Muscomorpha;  
CC OC Ephydroidea; Drosophilidae; Drosophila.  
CC OX NCBI\_TaxID=7227;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN-CANTON-S;  
CC RX MEDLINE-93165729; PubMed-8094559;  
CC RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;  
CC RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";  
CC RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).  
CC RN [2]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN-CANTON-S;  
CC RX MEDLINE-93165730; PubMed-8434015;  
CC RA Ballinger D.G., Xue N., Harshman K.D.;  
CC RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds  
CC calcium and contains a leucine zipper.";  
CC RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).  
CC CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO  
CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL  
CC OF CA+2 PER MOL OF PROTEIN.  
CC CC -1- SUBUNIT: HOMODIMER (PROBABLE).  
CC CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.  
CC CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF  
CC COMPOUND EYES AND OCCELLI.  
CC CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL  
CC DEVELOPMENT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L02111; AAA28405.1; -

Mon Apr 29 08:35:26 2002

EMBL: L05080; AAA28420.1; -  
 DR FlyBase; A47282; A47282; Cpn.  
 DR FlyBase; FBgn0010218; Cpn.  
 KW Calcium-binding.  
 FT CONFLICT 36 A -> AVAPAVVA (IN REF. 2).  
 FT CONFLICT 43 I -> T (IN REF. 2).  
 FT CONFLICT 64 I -> V (IN REF. 2).  
 FT CONFLICT 76 P -> A (IN REF. 2).  
 FT CONFLICT 100 P -> PP (IN REF. 2).  
 FT CONFLICT 126 VQ -> AP (IN REF. 2).  
 FT CONFLICT 154 I -> V (IN REF. 2).  
 FT CONFLICT 160 S -> T (IN REF. 2).  
 FT CONFLICT 160 A -> E (IN REF. 2).  
 FT CONFLICT 534 I -> T (IN REF. 2).  
 FT CONFLICT 699 V -> L (IN REF. 2).  
 FT CONFLICT 703 D -> E (IN REF. 2).  
 FT CONFLICT 721  
 SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;  
 Query Match 7.5%; Score 530; DB 1; Length 865;  
 Best Local Similarity 26.2%; pred. No. 2.4e-15;  
 Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;  
 QY 229 APTSKVLAKP-TPK---AETTKGALITPKPTTPKBPASTPKPTPKPTTIKSAPTT 284  
 DB 8 SPVSAPVAAPVTPSAVAAPVQVSPAAPV-----APAAPVATVPVAPPTTIASVQPAT 61  
 QY 285 PKPAPTITTKSAPTTP-KEPAPTITTKPAPTITTKPAPTITTKPAPTITTKSAPTTP 339  
 DB 62 VTIPAPAPIAASVTPVASVAPVVAAPT-----PAASPVSTPVVAQIPVASVAPV 116  
 QY 340 KEPAPTITTKKPAP-----TTKPEAPTITTKPAP-----TTKSA 437  
 DB 117 PVAATPTPVQIPVAAPVATPPVAASAPTTPAAVTPVISPVIASPPVVPVANTTPVPAAPV 176  
 QY 386 PREPA--PTAPKKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTK 437  
 DB 177 AAVPAPVAVPVLAPAVAPVAVVAPVETPAPVAPVETPAPVAPVETPAPVAPVETPAPVAPV 236  
 QY 438 PTTTKEP-----APTTKSAPTTPKESPTTKKEPA--PTTKPAPTITTKPAPTITTKPAPTITTK 493  
 DB 237 PLAAAPVVPVAPVATPTTPVAPAAASPHVSAPVAVETAVAPVSAVETPPVAAATLTTP 296  
 QY 494 TTP-----KEPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTK 534  
 DB 297 ETALAPVVAESQVAANTVATPPAPPEPAPVETPAPVAPVETPAPVAPVETPAPVAPVETPAPVAPV 356  
 QY 535 TPEKL--APTTPKEPAPTITTEELAPTTPPTTPEEP---APTTPKAAAPTTPKEPAPT 590  
 DB 357 AAESIPAPVATTPVPAT---LAVTDPDVASAVPELPVAPVAPVSAVETPPVAPVETPPVAPVETPPVAPV 413  
 QY 591 TPKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTK 639  
 DB 414 V-----LPPVAAEVPVAVAEETPETPAPASAPVIAALDIPVAPVIAAPSDAPAEAPSA 469  
 QY 640 LAPTTTKEPTSTSDKAPTTPKGTAPTTP-----KEPAPTITTKPAPTITTKPAPTITTKPAPTITTK 691  
 DB 470 AAPVSTTP--TTASVPTTAPAAVPTPEIDVSVLSEAAIETPPVAPPVEVETVAVADV 527  
 QY 692 TLKEPA-----PTTKKAPKAPLAPTTT---KGPTSTTSKAPAPT-----PKETAPT 737  
 DB 528 APPEAAADLIEPVEPPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 585  
 QY 738 PKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTKPAPTITTK 787  
 DB 586 EVAVAPITAPETIP---EPESLATPTEIPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 641  
 QY 788 SAETPKALENSKEPVPPTTKPAATKPEMTTAKDKITERDLRTP--ETTTAAPT 845  
 DB 642 TVE--FPEAVAEKVLDPAI--TEAPVTTQEPDVANINDGAPATEI--TTPAVEIVTAAEVS 697  
 QY 846 KETATTEKTESKITATTTQVST-----TTQDTTPPKITLTKTTLA--PKVTTTKKT 898

698 DIAIPVDIPVPQEIYAEIPETDTKPAEIVEQSTIPIEAPVPEVSKYAEPIVISEAPAA 757  
 899 ITTTEIMNKPEETAKPKDRATNSRATTPKPKOKPKAKPKATSTKPKTMRVRKPKTPT 958  
 758 EVPITAGNDPNTSVGISEVPTIAEKPEVEEPTSS--EIQSSSPS--DSVPVAKITPL 813  
 959 PRKMTSTMPENLPTSRIAEAM--LQTTTRPNTQNSKLVEV-----PKS 1001  
 814 LRDLQTTDVSLLAATAATLDAIGELKDKOKARNOQVMDRLCEIKILGPPKS 864

RESULT 11  
 FPL\_MYTGO STANDARD; PRT; 872 AA.  
 AC Q25434;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCFP1).  
 GN FPL  
 OS Mytilus coruscus (Sea mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=42192;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Foot;  
 RX MEDLINE=96394686; PubMed=8798340;  
 RA Haruya K.; Takeuchi Y.; Takeyama S.; Yamahara E.; Yamazaki F.; Odo S.;  
 RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and  
 its evolutionary implications";  
 RL J. Mol. Evol. 43:348-356(1996).  
 CC !- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
 PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
 ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
 FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
 CC !- SUBCELLULAR LOCATION: SECRETED.  
 CC !- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
 CC !- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
 CC !- PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND  
 ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL: D63777; BAA09850.1;  
 InterPro: IPR002964; Adhesive\_plaq.  
 InterPro: IPR002965; P-rich\_extensn.  
 PRINTS: PR01216; ADHESIVEI.  
 PRINTS: PR01217; PRICHEXTENS.  
 Signal: Repeat; Hydroxylation.  
 SIGNAL 1 20 POTENTIAL.  
 CHAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.  
 DOMAIN 21 41 NONREPEAT LINKER.  
 DOMAIN 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-  
 P-[PST]-[ST].  
 DOMAIN 184 192 NONAPEPTIDE 1.  
 DOMAIN 213 221 NONAPEPTIDE 2.  
 SEQUENCE 872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;  
 Query Match 7.3%; Score 518.5; DB 1; Length 872;  
 Best Local Similarity 29.2%; pred. No. 7e-15;  
 Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;  
 125 VVDEAGSLGNDGKVFVTPDTST--TQH-----NKVSTSPKITTAKPINRPSLPNS 175

```
Db 29 VYGSAYSGAGAYK-TLPGSHPGYSGKHVPVYKPMNKIPT-PYI--SKSPAPYKPKGY 84
QY 176 DTSKETSLVWNETTETETTTNKTQSTDGKEKTSKETSIAKDLAPTSKVL 235
Db 85 YPTKRYQPIYVGTNTYPPYKPIAKLSSYKAIKTTPYAKAKTSYPPSYK-----HKIT 139
QY 236 AKETPKAETTTKGPALTTKPTPTTPKPEASTTTPKEPTTTIKSAPTTKPEAPTTTK 294
Db 140 YPPTYK-----PKITYP--PTYKQKPSYPPSYKPKITYPTTYK-----PKITYPTTYK 185
QY 295 SAPT-TPKPEAPT--TTKPEAPTTPKPEAPTTPKPEAPTTPKSAPTTKPEAPT--BKK 349
Db 186 RPSYTPYKPKATYPTTKPKITYP-----PTYKRPSTY-----PYKPKTTPPYTKPKI 236
QY 350 PAPTTPKPEAPT--PREPTPTT--PKPEAPTTPKPEAPTTPKPEAPT-APKPEAPTTP 402
Db 237 SYPSYKPKASVSSYKSKTTPPYKPKISYPTTYKPKPSY-----PTYKPKVTYPTTY 292
QY 403 KEAPTTPKPEAPTTPKPSPTTPKPEAPTTPKSAPT-----TTKPEAPTTPKSAPTTPK 457
Db 293 K-PKPSYPTTYKPKITYPTT-YKPKPSYPTTYKQKPSYPTTYKSKSSYPTSYKSKKTYPP 350
QY 458 EPSPTTTKEPA-PTTPKPEAPTTPKPK-APT-TPK-----EPAPTTPKPEAPTTPK 507
Db 351 TYKPKITYPTTYKPKPSYPPSYKPKATYPTTYKPKITYPTTYKPKPSYPPSYKSKKTYPP 410
QY 508 APTAPKEAPTTPKETAP-----TTPKKLTPTTPEKLAPTTPKPEAPTTPPELAPTTPPE 562
Db 411 T-YKPKISYPTTYKPKASVSSYKSKTTPPYKPKISYPTTYKPKPSYPTTYKPKITYP 469
QY 563 PT-PTTPEAPTTPKKAAPTTPK-EPAPTTPKPEAPTTP-----KEPAPTTP--PKETAP 613
Db 470 PTYKPKPSYPTTYKPKITYPTTYKPKPSYPTTYKPKPSYPTTYKPKPSYPTTYKSKKTYPP 529
QY 614 TT--PKGTAPTTPKPEAPTTPKPKAPKELAPTTPKPTSTTSKAPAT-TPKGTAPTTPK 670
Db 530 PTYKPKITYPTTYK-----PKPSYPPSYKPKITYPTTYKPKITYPTTYKPKASYPPTYK 583
QY 671 EPAPTTPKPEAPTTPKGTAPTTPKPEAPTTP-----KKAPKELAPTTPKGTSTTSKDKPA 726
Db 584 -----PKITYPTTYK-PKPSYPTTYKQKPSYPTTYKSKSSYPTSYKSKKTY 628
QY 727 PTT--PKETAPTTPKPEAPTTPK--KPAPTTPET-----PPPTTSEVSTPTTKEPTT 775
Db 629 PPTYKPKITYPTTYK-PKPSYPPSYKPKITYPTTYKPKPSYPPSYKSKSSYPTSYKSKKTY 687
QY 776 IHKSPDESTELSAEPTPKALESPKPGVPTTKTPAATKPEMTTAKDKT-----TERDL 831
Db 688 Y---PPSYKPKKTYPTTYK-----PKISYPTTYK-----TKPSYPPSYKSKKTYPTTYKPK 735
QY 832 RTTPTTTAAAPKMTKETATTTKTTESKITATTTQVTTSTTQDTTTFKITTLKTTTLAPK 891
Db 736 ISVPSTYKAKP-----SYPTTYKPKPSYASSYKPKITYPTTYKPKPSYASSYK-----PK 785
QY 892 VTITTKTITTTMKNPEETAKPKDRATNSKATTPKPKQ---KPTKAPKPK-TSTKPKK-T 946
Db 786 I-----RYPPTYKPKP-----SYASSYKPKIRYPPTYKPKPSYASSYKPKKIT 827
QY 947 MPVRKPKKTT--PTPKMTSTMPKELNPTSRIAEAMLOTTTPRNO 988
Db 828 YPTTYKPKISYPTTYKPKITYPTTYKPKPSYPPSYKPKISYPSQ 871

RESULT 12
MUC1_XENLA
ID MUC1_XENLA STANDARD; PRT; 662 AA.
AC Q05049;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 01-OCT-1994 (Rel. 30, Last annotation update)
DE INTEGRATORY MUCIN C.1 (FIM-C.1) (FRAGMENT).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RX MEDLINE=93077556; PubMed=1447205;
RA Hauser F., Hoffmann W.;
RT "P-domains as shuffled cysteine-rich modules in integumentary mucin
RT C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic
RT polymorphism."
RL J. Biol. Chem. 267:24620-24624(1992).
CC -!- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: SKIN.
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -!- SIMILARITY: CONTAINS 6 P-TYPE (TREFOL) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L02115; AA074725.1; -
DR PIR: A45155; A45155.
DR HSP: P04002; IWFA.
DR InterPro: IPR000519; P_trefol.
DR Pfam: PF00088; trefol; 6.
DR SMART: SM00018; P; 6.
DR PROSITE: PS00025; P_TREFOL; 6.
KW Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
FT NON_TER 1
FT DOMAIN 81 144
FT 8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT ALA/THR-RICH.
FT REPEAT 81 88
FT 1-1.
FT REPEAT 89 96
FT 1-2.
FT REPEAT 97 104
FT 1-3.
FT REPEAT 105 112
FT 1-4.
FT REPEAT 113 120
FT 1-5.
FT REPEAT 121 128
FT 1-6.
FT REPEAT 129 136
FT 1-7.
FT REPEAT 137 144
FT 1-8.
FT DOMAIN 161 202
FT P-TYPE 1.
FT DOMAIN 218 301
FT 8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT REPEAT 218 224
FT 2-1.
FT REPEAT 225 239
FT 2-2.
FT REPEAT 240 249
FT 2-3.
FT REPEAT 250 259
FT 2-4.
FT REPEAT 260 275
FT 2-5.
FT REPEAT 276 287
FT 2-6.
FT REPEAT 288 294
FT 2-7.
FT REPEAT 295 301
FT 2-8.
FT DOMAIN 306 347
FT P-TYPE 2.
FT DOMAIN 353 394
FT P-TYPE 3.
FT DOMAIN 402 522
FT 12 X APPROXIMATE TANDEM REPEATS,
FT THR-RICH.
FT REPEAT 402 411
FT 3-1.
FT REPEAT 412 419
FT 3-2.
FT REPEAT 420 431
FT 3-3.
FT REPEAT 432 443
FT 3-4.
FT REPEAT 444 453
FT 3-5.
FT REPEAT 454 460
FT 3-6.
FT REPEAT 461 472
FT 3-7.
FT REPEAT 473 479
FT 3-8.
FT REPEAT 480 491
FT 3-9.
FT REPEAT 492 498
FT 3-10.
FT REPEAT 499 515
FT 3-11.
```

Db 437 TTTT-----TTKA-----TTTTTTTTTTTT--KAT-- 463  
 QY 831 LRTTETTTAAAPKMTKEATTEKTESKIATTTQVTTSTTTQDTTFKITTLLKTTTLLAP 890  
 Db 464 -TTTTTTTTT-----TTTTTKAT-----TTTTTTTTTTTTTTKATTTTTTTTTT 510  
 QY 891 KVTTKTKIITTEIMNKPEETAAPKDRATNSKATPPKQKPTKAPKPTSKKPK-----T 946  
 Db 511 TTTTTKATTTTTTSGCKME-----PSKRDCGYPGITESQCRSGCCFPDSSIPQTWKCFYS 566  
 QY 947 MPRVRKPKTTPTPR 960  
 Db 567 LPQVADCKVAPSSR 580

RESULT 13  
 RPBI\_HUMAN  
 ID RPBI\_HUMAN STANDARD; PRT; 1970 AA.  
 AC P24928;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPBI).  
 GN POLR2A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 ON NCBI\_taxid=9606;  
 RX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=92178992; PubMed=1542581;  
 RA Winterth M., Acker J., Vicaire S., Vigneron M., Keding C.;  
 RT "Complete sequence of the human RNA polymerase II largest subunit.";  
 RL Nucleic Acids Res. 20:910-910(1992).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=95347616; PubMed=7622068;  
 RA Mita K., Tsuji H., Morimyo M., Takahashi E., Neno M.,  
 RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;  
 RT "The human gene encoding the largest subunit of RNA polymerase II.";  
 RL Gene 159:285-286(1995).  
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- PTM: THE TANDEM 7 RESIDUE REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X63564; CAA45125.1; -  
 CC EMBL; X74874; CAA52862.1; -  
 CC EMBL; X74873; CAA52862.1; JOINED.  
 CC EMBL; X74872; CAA52862.1; JOINED.  
 CC EMBL; X74871; CAA52862.1; JOINED.  
 CC EMBL; X74870; CAA52862.1; JOINED.  
 CC PIR; S21054; S21054.  
 CC MIM; 180660; -





```
Db 243 KEOGEEEEEEGEGEEAATTS-----DPAEEAASP-----EKETKSPVKEEAKSPAEAKS 293
Qy 259 PTTTKEPA-STTPKEPTTTTIKSAPTTTKEPA---PTTTKSAPTTTKEPA-----PTTTK 309
Db 294 PAEAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 352
Qy 310 EP-----APTTTKEPA-----PTTTTKEPA-----PTTTTKEPA-----PTTTTKEPA-- 351
Db 353 SPGEAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 411
Qy 352 --PTTTTKEPA---PTTTTKEPA---PTTTTKEPA---PTTTTKEPA---PTTTTKEPA---PTA 393
Db 412 KSPVEAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 470
Qy 394 PKKPAPTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 444
Db 471 AKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 528
Qy 445 APTTTKS-----APTTTKEPSPTTTTKEPA-----PTTTTKEPAPTTTPKEPAPTTTPKEPAPTT 495
Db 529 VKEGAKSLAEAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPK 588
Qy 496 PKEPAPTTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 555
Db 589 AKPKEAKSPADIRSPQVKSAPKSPAEKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 644
Qy 556 APTTPEPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 615
Db 645 ---KAKEPPKVEEKTAPKTVEKSKKDEAPKEAQKP-KAEKEPLETEKPRDSPGEA 700
Qy 616 PKGTAPTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 675
Db 701 KKEEA-----KEKKAAPPEETPAKLGVEAKPKKEAEADAKA-----KEPSKP 744
Qy 676 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 735
Db 745 SEKE-----KEEVAAPPEKDTKE-----EKTESKKREKPKMEAK 784
Qy 736 TTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 782
Db 785 AKED-----KGLPOEPSKPKTEKAEKSSSTDQKDSQPSSEKAPED 824
```

Search completed: April 26, 2002, 16:33:31  
Job time: 641 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:26:24 ; Search time 61.21 Seconds  
(without alignments)  
481.977 Million cell updates/sec

Title: AA8

Perfect score: 7064

Sequence: 1 MAWKTLPIYLLLSVFVIQ.....ARATTRSGQTLKVVWYVNC 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	7.4	805	US-09-103-429A-4	Sequence 4, Appl
2	508	7.2	1837	US-08-928-361B-5	Sequence 5, Appl
3	506.5	7.2	744	5202236-25	Patent No. 5202236
4	498.5	7.1	786	US-09-103-429A-3	Sequence 3, Appl
5	489	6.9	1721	US-08-700-651-5	Sequence 5, Appl
6	489	6.9	1721	US-08-928-361B-6	Sequence 6, Appl
7	488.5	6.9	826	US-07-638-431-2	Sequence 2, Appl
8	488.5	6.9	826	PCT-US92-00018-2	Patent No. 5202236
9	476.5	6.7	652	5202236-13	Sequence 5, Appl
10	452	6.4	960	US-09-219-849-5	Sequence 5, Appl
11	424.5	6.0	1867	US-08-479-537A-5	Sequence 5, Appl
12	424.5	6.0	1867	US-09-083-116-5	Sequence 5, Appl
13	424.5	6.0	2035	US-08-479-537A-2	Sequence 2, Appl
14	424.5	6.0	2035	US-09-083-116-2	Sequence 2, Appl
15	419.5	5.9	2476	US-08-276-967-2	Sequence 2, Appl
16	419	5.9	1185	US-09-041-886-23	Sequence 23, Appl
17	417	5.9	829	US-08-642-255-132	Sequence 132, Appl
18	417	5.9	829	US-08-397-633A-53	Sequence 53, Appl
19	417	5.9	837	US-08-175-155-68	Sequence 68, Appl
20	417	5.9	837	US-08-477-509B-103	Sequence 103, Appl
21	417	5.9	837	US-08-642-255-101	Sequence 101, Appl
22	417	5.9	837	US-08-707-237A-75	Sequence 75, Appl
23	417	5.9	837	US-08-482-085B-103	Sequence 103, Appl
24	417	5.9	897	US-08-397-633A-50	Sequence 50, Appl
25	413.5	5.9	907	US-08-783-774-2	Sequence 2, Appl
26	413.5	5.9	907	PCT-US95-04611A-19	Sequence 19, Appl
27	404.5	5.7	408	US-07-609-716-65	Sequence 65, Appl

28 404.5 5.7 408 4 US-08-475-411A-65 Sequence 65, Appl

29 404.5 5.7 408 4 US-08-478-029A-65 Sequence 65, Appl

30 398.5 5.6 682 1 US-08-642-255-126 Sequence 126, Appl

31 398.5 5.4 682 1 US-08-397-633A-36 Sequence 36, Appl

32 381 5.4 1848 4 US-08-296-791-6 Sequence 6, Appl

33 381 5.4 1848 5 PCT-US95-10661A-6 Sequence 6, Appl

34 376 5.3 960 4 US-09-219-849-6 Sequence 6, Appl

35 375.5 5.3 1231 3 US-08-904-263A-4 Sequence 4, Appl

36 374.5 5.3 1537 1 US-08-325-267A-2 Sequence 2, Appl

37 371 5.3 761 2 US-08-707-237A-84 Sequence 84, Appl

38 371 5.3 762 1 US-08-642-255-114 Sequence 114, Appl

39 371 5.3 762 1 US-08-397-633A-26 Sequence 26, Appl

40 370.5 5.2 1064 1 US-08-642-255-62 Sequence 62, Appl

41 368 5.2 762 1 US-08-642-255-120 Sequence 120, Appl

42 368 5.2 762 1 US-08-397-633A-31 Sequence 31, Appl

43 367.5 5.2 1187 1 US-08-320-559-28 Sequence 28, Appl

44 367.5 5.2 1187 3 US-08-545-860D-28 Sequence 28, Appl

45 367.5 5.2 1187 5 PCT-US94-04496-28 Sequence 28, Appl

#### ALIGNMENTS

RESULT 1

US-09-103-429A-4

; Sequence 4, Application US/09103429A

; Patent No. 6187558

; GENERAL INFORMATION:

; APPLICANT: Granados, Robert R

; APPLICANT: Wang, ping

; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin

; TITLE OF INVENTION: CDNA and Related Products and Methods

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.

; STREET: 118 No. 6187558th Tioga

; CITY: Ithaca

; STATE: NY

; COUNTRY: USA

; ZIP: 14850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09103,429A

FILING DATE: 24-JUN-1998

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Michaels, Christopher A

REGISTRATION NUMBER: 34,390

REFERENCE/DOCKET NUMBER: BTI-39

TELECOMMUNICATION INFORMATION:

TELEPHONE: (607) 256-2000

TELEFAX: (607) 256-3628

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 805 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Trichoplusia ni

TISSUE TYPE: peritrophic membrane

US-09-103-429A-4

Query Match 7.4%; Score 522; DB 4; Length 805;

Best Local Similarity 27.8%; Pred. No. 1.5e-25;



QY 518 -----TTPKETA-PTTP-----KKL 531  
Db 842 AGQIADTSLNLPVQTHKSTGLPIDWVGLPFDPKSGNLVHYPTNOTMSGLSVSLAAKNL 901  
QY 532 T-----PTTPEKLIAPTPE-----545  
Db 902 TVDTEYGLPIDTILTYPLDPVSLIPENPETGELFIDISDEIMNGTIAGIVSGISASES 961  
QY 546 ----KPAPTTPEE-----LAPIT-----PEEPTPTP--EEPAPTTPKAAA 580  
Db 962 LLSQKAPIDATNMVWGEFGLLNPAATGVMIPGSLGSEQTFFSEIDGGIIPPEVAA 1021  
QY 581 PNTKPEAPTTPKPEAPTTPE-----604  
Db 1022 ANADKFLSLPSPESIPEDQKIDISSELMYDIESRLIGOVSKRPIGSIAGDLNPI 1081  
QY 605 PTTKETAATTPKGTATTLKPEAPTTPKKAPKAPKELAPTTTKESTTSKAPAPTTPKGT 664  
Db 1082 MKTPTQDSVTKPIDPTT---GLFPNP--PTGHLINPTNNMTDSSFAGAYKVAVNGI 1136  
QY 665 APTT----PKPEAPTTPKGTAPTTLKPEAPTT-----KPAKP-- 707  
Db 1137 KTDNVYGLPVDIEITGLPKDPVSDIPFNSTTGELVDPSTGKPIPNNTAGIVSGKGLPIE 1196  
QY 708 ----ELAPTTTKGP-----TSTT-----SDKPAPTTPKETAPTTLKPEA--- 742  
Db 1197 DENGNLFPSTKPIDGNLNNLNPENSTVSGSTGSKPKPGIPVNGGVVDPDEAKDQ 1256  
QY 743 -----PTPKKAPATTPETPTTSEVSTPTTKE--PTTIHKS---DDESTPEL 787  
Db 1257 ADKKGDLVPTNSINKDPVNTQYSGNTGNINPETGKVIPLGSLPGLNPSFNTPQ 1316  
QY 788 SAETPKALE---NSPKPEGVPTTKTAPATKPEMTTAKDKTTERDLRTTPTTTAPKM 844  
Db 1317 TDEITGKPDVTVGLPDPSTGEIIDATKLPVGSVAGDEILTEVLNITDEVGLP-I 1375  
QY 845 TKEATTTETKESKITATTTQVSTTTQDTTFFKITTLKTLAPKVVTTKKTITTE- 903  
Db 1376 DLETGLRDP-----VSLGQLPNGTLVD-----PSNKKPIPGSHS 1411  
QY 904 --INNKPEETAAPKDRATNSKATTPKPKP--TKAP--KPTSTKPKKTMVRKPKT--- 955  
Db 1412 GFINGTSGEQSHKDPSTG-----KPLDNTGLPDEDSGLNINPETGDKLOGSHSGTF 1465  
QY 956 TTPPKMTSTWPELNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPK--SEDA----- 1004  
Db 1466 MPVPGKPGQGGIMTPEQILEAL-----NKLPTSNEVNISPRSSDAVDPDRPTNWW 1518  
QY 1005 -----GGAGETPHMLL-----RPHVFMEVTPMDYLPRVDPNOGII 1042  
Db 1519 NKISGQTFQVDGKKTIPGSAASVIHTALGTPQTQDPTTGLP--SDPSTGLPFTPGFNVLV 1576  
QY 1043 NPMLSD-----ETNIC-----NGKPDVGLTTL---RNGTLV-AF 1072  
Db 1577 DPOTGEQIKGVPVSVLYVKEKNIVTEAAAYGLPVDPKTGFPIDPISYLPFAKNCELIDPI 1636  
QY 1073 RGHVF 1077  
Db 1637 SGKIF 1641

RESULT 3

5202236-25

; Patent No. 5202236

; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,

; SUSAN L.; MCCANDLIS, RUSS; WEI, TENA; FILPUILA, DAVID

; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE

; PROTEIN

; NUMBER OF SEQUENCES: 39

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/528,762

; FILING DATE: 25-MAY-1990

; APPLICATION NUMBER: 82,456

; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945  
; FILING DATE: 24-NOV-1986  
; APPLICATION NUMBER: 650,128  
; FILING DATE: 13-SEP-1984  
; SEQ ID NO:25: 744  
; LENGTH: 744  
; 5202236-25

Query Match 7.2%; Score 506.5; DB 6; Length 744;  
Best Local Similarity 29.9%; Pred. No. 1.3e-24;

Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 157 PKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNTKOTSTDGCKEKTSAKE 216  
Db 24 PKMTYPTTKPKPSYPP-----TYKSKPTY-----KRIIT----- 53  
QY 217 TQSIKTSADKLAPTSKVLAKPTPKAETTTKGALTTTPKEPTTTPKE-----PASTTPKE 272  
Db 54 -----YPTTK--AKPS-----YPTTKPKKTYDPYKPKLTYDPYKPKP 92  
QY 273 PTPTTIKSAPT--TPKEPAPTTTKSAPTTPKEPAPTTTKPEAPTTTKPEAPTTTKEA--P 329  
Db 93 SYPTTKSKPTYKPKIYPTTKYAKPSYPTTKPKKTYPPPT--YKPKLTYPTTKPKASYP 151  
QY 330 TTKKSAPTTPKEPAPTTTPKPKAPT--TPKEPAPTT-----PKEPTTTPKEPAPTTKEP--- 381  
Db 152 PTYKPKPSYPP--PSYKTKTKTYPTTKPKLTYPTTKPKPSYPPSYKPKKTYPTTKPKLT 209  
QY 382 -APTTLKPEA--PTAPKKA--PTTKPEA--PTTKPEA--PTTKPEA--PTTKPEA 430  
Db 210 YPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKPS 269  
QY 431 -PTTKSAPTTPKEPAPTTTKSAPTTPKPSPTTKPEA--PTTKPEA-----PTTKPKP 483  
Db 270 YPSTYKAKPSYT---YPTTKYAKPTY---PTTKYAKPSYPTTKYAKPSYPTTKYAKP 322  
QY 484 A--PTTKPEA--PTTKPEA--PTTKKKA--PTAPKEPAPTTTPKETAPTTPKLTPTT 535  
Db 323 SYPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYKPKTYPS 382  
QY 536 PEKLAPTPE--KPAPTTPEELAPTTPPEPT--PTTKPEA--PTTKPA--AAPNT---P 584  
Db 383 TYKAKPSYPTTKYAKPSYPP---PTTKYAKPSYPTTKYAKPSYPTTKYAKPTYPSTYKAKP 438  
QY 585 KEPAPTTPKEPAPTT-----PKEPAPTTTPKETAPT--PKGTAPTTLKEPAPTTPKKPAPK 638  
Db 439 SYPTTKYKPKISYPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKP 498  
QY 639 ELAPTTPKEPTSTSDKPAPTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEA- 697  
Db 499 SYPTTKYKAPPTYPSTYKAKPTYK--ARPSYPP---PTTKYAKPS-----YPTTKYAKPSY 546  
QY 698 -PTTKPKKAPKELAPTTTKGPT--STTSKPAPTTPKETAPTTPKEPAPTTTKPKPA--PT 752  
Db 547 PPTTKYAKPTYK--AKPTYKAKPTYPTTKYAKPSYPTTKYAKPSYPP---PTTKYAKPSYPT 601  
QY 753 ---TPETPTTSEVSTPTTKE-----PTTIHKSPP-----ESTPELSAEP--PKALENS 799  
Db 602 YKAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKPSYPTTKYAKP 661  
QY 800 KPEPGVPTTKTAA-----TKPEMTTAKDKTTERDLRTTPTTTAAAP-----KMT 845  
Db 662 PSYP--PTTKYAKPSYPTTKYAKPSYPTTKYAKPTYKAKPTYKAKPTTPSTYKAKPSYPTTKHAKPT 719  
QY 846 KETATTTETKESKITATTT 865  
Db 720 YKAKPTYPSTYKAKPTYDPT 739

RESULT 4

US-09-103-429A-3

Sequence 3, Application US/09103429A  
Patent No. 6187558  
GENERAL INFORMATION:  
APPLICANT: Granados, Robert R  
APPLICANT: Wang, Ping  
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
TITLE OF INVENTION: cDNA and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
STREET: 118 No. 6187558th Tlaga  
CITY: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/103,429A  
FILING DATE: 24-JUN-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 786 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-103-429A-3

Query Match 7.1%; Score 498.5; DB 4; Length 786;  
Best Local Similarity 29.1%; Pred. No. 4.6e-24;  
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

QY 238 PTPKAETTKGPAITTPKEPTTPKPEASTTPKEPTTTIKSAPTTPKPEAPTTPKTSAP 297  
DB 93 PGPAE-TTQAPAT- - - - -QAPTTOAPTIT- - - - -QAPTITQ- - - - -ATTITQAP 135  
QY 298 TTPKEAPTTPKPEAPTTPKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKPE 357  
DB 136 TTQ-APTITQ- - - - -QAPTITQ-APTITQAPTITQ-APTITQ-APTITQ-APTITQ- 183  
QY 358 PAPTTPKEPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 414  
DB 184 -APTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQ 241  
QY 415 - - - - -PTTKEPSPTTPK- 427  
DB 242 ELLPNCRADEDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLFNPVQRCDSANVEC 301  
QY 428 - - - - -EPAPTTT- 434  
DB 302 DGEISPAFVTEGNEDEIDIGLLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERR 361  
QY 435 - - - - -KSAPTTPKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 467  
DB 362 CGAGTHFSFELQOCDHIELVGLTLPGBSEEDVDEDA-CTGWYCPTEPIEWEPLNCPG 420

QY 468 APTT- - - - -PKEP- - - - -APTTPKKDAPT- - - - - 487  
DB 421 ADFSLHLLPHESDCGOYLQCVHGOTIARPCGNLHFSPATQSCSPVTAGCQVFECDS 480  
QY 488 - - - - -PKEPAPTTPKPEAPTTPKPAAPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 544  
DB 481 NOCTSTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 533  
QY 545 EKPAPTPEELAPTTPEEPTTPPEAPTTPKAAAPT- - -PKEPAPTTPKPEAPTTPKE 602  
DB 534 - - - - -PAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 587  
QY 603 PAPTTPKETAPTTPKGTAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 662  
DB 588 AAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 632  
QY 663 GTAPTTPKPEAPTTPKPEAPTTPKGTAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 722  
DB 633 TAAPT- - - - -AAPAPTTPVTPPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 677  
QY 723 DKPAPTTPKETAPTTPKPEAPTTPKPKP 749  
DB 678 - - - - -APATTPEDDDIDP- - - - -PLPNDPINP 699

RESULT 5  
US-08-700-651-5  
Sequence 5, Application US/08700651B  
Patent No. 6015882  
GENERAL INFORMATION:  
APPLICANT: PETERSEN, CAROLYN  
APPLICANT: LEECH, JAMES  
APPLICANT: NELSON, RICHARD, C.  
APPLICANT: GUT, JIRI  
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
FILE REFERENCE: 480.19-4 (HV)  
CURRENT APPLICATION NUMBER: US/08/700,651B  
CURRENT FILING DATE: 1997-08-14  
EARLIER APPLICATION NUMBER: 08/415,751  
EARLIER FILING DATE: 1995-04-03  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 1721  
TYPE: PRT  
ORGANISM: Cryptosporidium parvum  
US-08-700-651-5

Query Match 6.9%; Score 489; DB 3; Length 1721;  
Best Local Similarity 22.1%; Pred. No. 4.5e-23;  
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

QY 111 KKNRTKKKPTKPPVDDEAGSLDNGDFKVTTPDTSTQHNVKYSTPKITAKPINRPS 170  
DB 116 RSNETK- - - - -TPEPSANTYAGYVRSN-ETKTEPSANT- - - - -NFLVDPKI- - - - -N 158  
QY 171 LPPNSDT/SKE- - - - -TSITVNETTIVETKETTITNTKOTSTDGKEKTTSAK 215  
DB 159 APCNSESEGOQIFDMGSKVYIPYTKCVGKHHTTTTTTTTTTTTTTTTTTTTTT 208  
QY 216 ETQSIKTSKADLAPT SKVLAKPTPKAETTTKGPALTTPKPEPTTPKPEASTTPKPEPT 275  
DB 209 - - - - -TT 240  
QY 276 TTIKSAPTTPKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKSA 335  
DB 241 TT 300  
QY 336 PTTTPKPEAPTTPKPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTAPK 395



```
QY 456 PKE-----PSPTT----- 463
Db 407 TSETESVKEWLEKNGEAKGATYVYVIGKDGRIENGMAFTMIPNDHDFVRFK 466
QY 464 -----TKEP-----APTTPKE 474
Db 467 VKDVGNTISVRCRGAGKLEFPDRSLDFTIPPVAGHNSCSIIYVSGDGIHVSYPGSKD 526
QY 475 PAPTTPKKPAPTTPKE-----PAPTTPKEPAPTTPKKPAPTAPKEPA 516
Db 527 VSLIS-----APIQSELFNEVYDCTAKYGAHSGYQTSADFVTTTAKPTTTT-TGA 580
QY 517 PTPPKETAPTTPKKLPTTPTEKLAPTPEKPAPTPEELAPTPEPTTPPEAPTTP 576
Db 581 PGQPTTTTSGSPSKPTTTTAKATTT-----TTLNIITTTTOKPTTTT-----TT- 627
QY 577 KAAAPNPKPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPA 636
Db 628 --KVPKPPATTTTLKPIVTTTAKATTTTIVPTT-----TTTTKRDEMTTTTTL 680
QY 637 PK-----ELAPTTPKEPTS-----TTSDKPAPTTPKG--TAPTTPKEPAPT 676
Db 681 PDIGDIEITPIEKMLDKYRMIYDYNGLLSDNDEPIPGSQAGIADTSNLFVQTH 740
QY 677 PKEPAPTTPKGTAPTTLKE--PAPTTPKKPAKELAPTTPKGTSTTSKPAPTTPKET 733
Db 741 KSTGLPIDPMVGLPFDPKSNLVHPYTNQTMGLSVSYLAAKNLVDTDETYG--LPIDT 798
QY 734 APTTPKEPAPTTPKKPAPTTPETPTTSEVSTPT-----TTKEPTTIHKS----- 779
Db 799 LTGYPLDPSVLIPFN--PETGELFDPISDEIMNGTIAGIVSGISASESLLSOKSALIDPA 856
QY 780 -----PDEST-----PELSAETPKALENSPK 801
Db 857 TMVYGEFGLLNATGVMIPGLGSEPTQFSPEDIGIIPPEVAAAANADKFKUSIP- 915
QY 802 EPGVPTTKTPAATKPEMTTAKDKTTER-----DLRTTPTTTTAAPKM 844
Db 916 -PSVP-----ESIP-----KQKIDISISELMYDIESGRIGQVSKRPIPGSIAGDLNP 963
QY 845 TKETATTEKTSKITATT--TQVSTTTQDTPFKITTLKTTTLAKPVTTTKTITTE 903
Db 964 IMKTPTQDTSVTKPIDPTTGLFPNPTTGLINPTNNNTMDSSFAGAYKAYVNSGIKTDN 1023
QY 904 IMNKP--EETAKPKD-----RATNSKATTPKPKPTKAPKPTSTKPKTMP----- 948
Db 1024 YVGLPVEITGLPKDPSDIPFNSITGELVDPSTGKFINNSTAGIVSGKGLPIEDENG 1083
QY 949 -----RYRKPKTTPTPKMTS--TMPE----- 968
Db 1084 NLEDPSTNLPIDGNQVLNVPETNSVSGTSGTKPKPGIPVNGGVVVPDEAKDQADKG 1143
QY 969 -----LNPSTRIABMLQTTTRPQNTNSKLIVFNPKSEDAGAGETPHMLLRPHVFP 1023
Db 1144 KGLIVPPTNSINKDPVTNTQYSNTTGN-----INP--ETGKIVPGLPSGLNVPSPNTP 1197
QY 1024 EVTPD-----MDYLPRYPNOGIILNP-----MLSDETNICNGK----- 1056
Db 1198 QQTDEITGKPVDTVGLPYDPSGTEIIDPATKLPVGSVAGDEILTEVLNITDDEVTLGP 1257
QY 1057 -----PVDGLATTLRNGTLV 1070
Db 1258 IDLETGLPDPVPSGLPLPNGTLV 1281
```

## RESULT 7

```
US-07-638-431-2
; Sequence 2, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
```

```
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumsmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; NUMBER OF INVENTIONS: immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-638-431-2
```

Query Match 6.9%; Score 488.5; DB 1; Length 826;

Best Local Similarity 26.4%; Pred. No. 2.1e-23;

Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

```
QY 274 TPTTIKSAPTTPKEPA-PTTTKSAPTTPKEPA-PTTTTKEP-----APTTPKEP-APTTPKE 326
Db 273 TPCKVRCPOIPVPIVINKIPEKPSNPEEPVNDNDNNNNNNNNNNNNNNNNNN 332
QY 327 PAPTTPKSAPTTPKEP-APTTPKKP-APTTPKEP-APTTPKEP-APTTPKEP-APTTPKEP 382
Db 333 PNNPNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNDPSNPNHPK 392
QY 383 PTPPKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTT 440
Db 393 RRNPKRNNKPKPKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPN 443
QY 441 TKEPAPTTPKSAPTTPKEPSPTTTKEPAPTTPKEPA-PTTPKKP-APTTPKEPAPTTPKE 498
Db 444 PKNPNNP-----NEPSNPKPNP-----EPNPNPEPSNPNPEPSNPNPEPSNPNPEPSNPN 492
QY 499 PAPTTPKKKPAPTAPKEP-APTTPKETAPTTPKKLTPTTPEKLAPTTPKEPAPTTPPEELAP 557
Db 493 PS-----NPNEPSNPNPEPSNPNPEPSNPNKPNP-----EPSNPNKPNP-----NEPSN 523
QY 558 TTPPEPTTPPEEPAPTTPKAAAPNTTPKEP-APTTPKEPAPTTPKEPAPTTPKETAPTTP 616
Db 524 SNPNE--PLNPNEP-----SNPNEPSNPNPEPSNPNPEPSNPNPEPSNPNPEPSNPNPEPSNPN 564
QY 617 KGTAAPTTLKEPAPTTPKKPAKELAPTTPKPTTSDKPAPTTPKGTAPTTPKEPAPTTP 676
Db 565 -----PSNPEEPNPEE--PSNPKPEP-----SNPEEPINPEELNPKPSNPEESN 606
QY 677 PKEPAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTPKGTSTTSKPAPTTPKETAPTTP 736
Db 607 PKEPI-----NPEESNPKPEP-----INPEDNENPLIIQDEPIERDNSNVIPI 649
```









STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,116  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/479,537  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1867 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1727  
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."  
OTHER INFORMATION:  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."  
OTHER INFORMATION:  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA or ACG; and Asn = AAT or AAC."  
OTHER INFORMATION:  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."  
OTHER INFORMATION:  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."  
OTHER INFORMATION:  
US-09-083-116-5

Query Match 6.0%; Score 424.5; DB 4; Length 1867;  
Best Local Similarity 23.7%; Pred. No. 6e-19;  
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;  
QY 181 TSLTVNKEITVETKETTTTNKQITSDGKEKTTSAKETQSTKTSKAD-LAPTSTKVLAKPT 239  
Db 16 TVLTV-----VTGSHASSTPGGEKETSATORSSVPSSTERNNAVSWTSSVLSHS 65  
QY 240 P-KAETTTTGP--ALTTTPKEP-----TPTTPKEPASTTPKEPTTTTIKSA 282  
Db 66 PGSGSSTTQGDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTTPPAHDTV--SAP 122  
QY 283 TTPKEPAPTTT-----KSAPTTTPKEP-----APTTPKEPAPTTTPKEPAPT 323  
Db 123 --DNKPAPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 180  
QY 324 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KPAPTTTPKEPAPTTPKEPTPTT 370  
Db 181 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 240  
QY 371 PKE--PAPTTTPKEP-----APTTPKEPAPTAPK-----KPAPTTTPKEPAPTTPKEPAPT 417  
Db 241 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 300  
QY 418 TKPSPTTPKEPAPT--TKSAPTTTTPKEPAPT-----TKSAPTTTPKEPSPTTK-----E 466  
Db 301 APDXRP--XPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 359  
QY 467 PAPTTPKEPAPTTPK-----KPAPTTTPKEPAPTTPK-----EPAPTTTTPKEPAPTTPKEPAPT 518  
Db 360 SAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 419  
QY 519 TPKETAP-----TTPPKLTPTTTPKEKAPTTTPKEPAPTTPPELAPTTTPPEPTTTPPEEAPT 574  
Db 420 SAPDXRXPXGSTAPXAHGVT-----APDXRXPXGSTAPXAHGVTAPDXRXP-----PGST 471  
QY 575 TPKA---AAPNTTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGT--APTTLKEPAP 629  
Db 472 APXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTAPDXRXPXPGS 530  
QY 630 TTP-----KKPAPKELAPTTTTPKEPTSTTSDKAP--TTPKGTAPTTPKEPAPTTP 677  
Db 531 TAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDXRXPXGSTAPXAHGVT-----APDXR 585  
QY 678 KEPAPTTTPKG-----TAPTTLKEPAPTTP-----KKPAPKELAPTTTTPKEPTSTTSD 723  
Db 586 PPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDX 644  
QY 724 KPAP--TTPK-----ETAPTTPKEPAPTTP-----KKPAPTTTPPETTTPPTTSEVSTP 767  
Db 645 RPXPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXPGS--TAPXAHGVTAP 702  
QY 768 TTTTKEPT-----IHKSPDESTPELSAPTPKALENSPKPEPGVPTTKTPAA----- 813  
Db 703 DXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAP 762  
QY 814 -TKPEMTTAKDK---TTERDLRTTPETTTAAAPKMTKATATTTTTEKTSKITATTTQVTS 869  
Db 763 DXRXPXGSTAPXAHGVTAPDXRXPXGST--APXAHGVTAPDXRXPXGSTAPXAHGVT 820  
QY 870 TTTQDTPPFKITTLKTTTLAPKVTTTKTTTITTEIMNKPEETAKPKDRATNSKATTPKQ 929  
Db 821 APDXRXPXPG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTAPDXRXP 868  
QY 930 KPTKAPKKPTSTKKPKTTPRVKPKTTPTPKMTSTMPELNPTSRIAEAMLQTTTRPNOT 989  
Db 869 -GSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTSA--PDXRXP-----PGST 911  
QY 990 -PNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPDMXYLPRVFNOCIIINPMLSD 1048  
Db 912 APXAHGVTAPDXRXPXGSTAPXAH-----GVTAPDXRXPXGSTA---PXAHG 957  
QY 1049 ETNICNCKPVDGLTT-LRNGTLVAFRGHYFWMLSPFPSPARRITETWVGIPSPID 1103

Db 958 VTSAPDXRXPXGSTAPXAHGVTSAPD-----PDXRXPXGSTAPXAHGVTSAPD 1003

RESULT 13

US-08-479-537A-2

Sequence 2, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMBER, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREUVENI, Mara

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1899

OTHER INFORMATION: /note= "The amino acids spanning

OTHER INFORMATION: 128 to 1899 constitute a repeated

OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such

OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa

OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note= "Amino acids 1 to 21 are a

OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 6.0%; Score 424.5; DB 2; Length 2035;

Best Local Similarity 23.7%; Pred. No. 6.6e-19;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 181 TSLTVNKETTVEKTTTNNKQSTDCGKETTSAKETQSIKTSKAD-LAPTSKVLAKPT 239

Db 16 TVLTV-----VTGSGHASSTPGCEKETSATQSSVPSSTERNKAVSMTSSVLSHS 65

QY 240 P-KAETTTKGP--ALATPKP-----TPTTPKEPASTTPKEPTTPTTKSAP 282

Db 66 PGSGSSTTQGGDVLAPATEPASGAATWGQDVTSPVTRPALGSIPTPAHDVT---SAP 122

QY 283 TTPKEPAPTTT-----KSAPTTPKP-----APTTTKEPAPTTTKEPAPTT 323

Db 123 --DNKPAGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 180

QY 324 TKE--PAPTTT-----KSAPTTPKPAPTTPK-----KPAPTTPKPAPTTTKEPTPTT 370

Db 181 APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 240

QY 371 PKE--PAPTTKEP-----APTTTKEPAPTPK-----KPAPTTPKPAPTTTKEPAPTT 417

Db 241 APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 300

QY 418 TKEPSTTPKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTKEPSPPTTK-----E 466

Db 301 APDXRP--XPGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 359

QY 467 PAPTTTKEPAPTTPK-----KPAPTTTKEPAPTTPK-----EPAPTTTKEPAPTTKEPAPTT 518

Db 360 SAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 419

QY 519 TPKETAP-----TTPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPPTTPEEPAPT 574

Db 420 SAPDXRXPXGSTAPXAHGVTS-----APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 471

QY 575 TPKA-----AAPNTPKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 629

Db 472 APXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 530

QY 630 TTP-----KKPAPKELAPTTTKEPTSTTSKAPAP--TTPKGTAPTTTKEPAPTTT 677

Db 531 TAPXAHGVTSAPDXRXPXGSTAP--XAHGVTSAPDXRXPXGSTAPXAHGVTS-----APDXR 585

QY 678 KEPAPTTPKG-----TAPTTTKEPAPTT-----KKPAPKELAPTTTKEPTSTTS 723

Db 586 XPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 644

QY 724 KPAP--TTPK-----ETAPTTTKEPAPTT-----KKPAPTTPEPTTSEVSTP 767

Db 645 RXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAP 702

QY 768 TTTTKEPT-----THKSPDESTPELSAETPKALENSPKPEKPVPTTKTAA-----813

Db 703 DXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAP 762

QY 814 -TRPEMTTAKDK----TTERDLRTTPTTAAAPKMTKETATTTKTESKITATTTQVTS 869  
Db 763 DXRPXPGSTAPXAHGVTSPDXRPGST--APXAHGVTSPDXRPGSTAPXAHGVT 820  
QY 870 TTTQDTTPPKITLTKTTTLAPKVTTTKKITTITEIMNKPEETAAPKDRATNSKATTPKQ 929  
Db 821 APDXRPGXPG-----STAPXAHGVT-----SAPDXRPGXPGSTAPXAHGVTSPDXRPGX- 868  
QY 930 KPTKAPKPTSTKPKTMRVRPKTPTPRKMTSTMPNLNPTSRTAEAMLQTTTRPNQT 989  
Db 869 -GSTAPXAHGVTSPDXRPGXPGSTAPXAHGVTSA-PDXRPGXPGSTAPXAHGVTSPDXRPGX- 911  
QY 990 -PNSKLVENPKSEDAGGABGETPHMLLRHVFEMPEVTDMYLPVPGNOGIIINPMLSD 1048  
Db 912 APXAHGVTSPDXRPGXPGSTAPXAHGVTSA-----GVTSAPDXRPGXPGSTAPXAHGVT 957  
QY 1049 ETNICNKPVDGLTT-LRNQTLVAFRGHYFWMLSPPSPSPARRITEVNGIPIPID 1103  
Db 958 VTSAPDXRPGXPGSTAPXAHGVTSA-----PDXRPGXPGSTAPXAHGVTSPDXRPGX- 1003

## RESULT 14

US-09-083-116-2

Sequence 2, Application US/09083116

Patent No. 6203795

GENERAL INFORMATION:

APPLICANT: CHAMON, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREUVENI, Mara

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER &amp; MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/083,116

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/479,537

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1899  
OTHER INFORMATION: /note= "The amino acids spanning  
128 to 1899 constitute a repeated region wherein the repea  
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
OTHER INFORMATION: repeats varies from 1 to 40."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
OTHER INFORMATION: Xaa Xaa which is the codon for pro or Ala wherein pro = CC  
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,  
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
OTHER INFORMATION: which is the codon for pro or Ala wherein pro = CCT, CCC,  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
OTHER INFORMATION: 21 amino acid precursor sequence."  
US-09-083-116-2

Query Match 6.0%; Score 424.5; DB 4; Length 2035;

Best Local Similarity 23.7%; Pred. No. 6.6e-19;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 181 TSLTVNKETTETKTTTNTKQSTDKGKTTTSAKETSQTSKAKD-LAPTSTKVLAKPT 239  
Db 16 TVLTV-----VTGSGHASTPGGKETSATQSSVPSSTKNAVMTSSVLSHS 65  
QY 240 P-KAETTTKGP--ALTTTPKEP-----TPTTPKEPASTTPKEPTTTIKSAP 282  
Db 66 PGSGSSTQGDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTTPPAHDVT---SAP 122  
QY 283 TTPKEPAPTTT-----KSAPTTTPKEP-----APTTPKEPAPTTTPKEPAPT 323  
Db 123 --DNKPAPGSTAPXAHGVTSPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVT 180  
QY 324 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KPAPTTPKEPAPTTPKEPAPT 370  
Db 181 APDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVT 240  
QY 371 PKE--PAPTTKEP-----APTTPKEPAPTAPK-----KPAPTTPKEPAPTTPKEPAPT 417  
Db 241 APDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVT 300  
QY 418 TKEPSPTTPKEPAPT--TKSAPTTTTPKEPAPT-----TKSAPTTTPKEPSPTTTPK----E 466  
Db 301 APDXRP-XPGSTAPXAHGVTSPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVT 359  
QY 467 PAPTTPKEPAPTTPK-----KPAPTTPKEPAPTTPK-----EPAPTTPKKTAPKAPAPT 518  
Db 360 SAPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVT 419  
QY 519 TPKETAP-----TTPKKLTPTTPEKLAPTTPKEPAPTPEELAPTTPEPTPTTPEEAPT 574  
Db 420 SAPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVTSPDXRPGSTAPXAHGVT 471  
QY 575 TPKA----AAPNTTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGT-APTTLKEPAP 629

Db	472	APXAHGVT	SAPDXRXPXG	STAPXAHGVT	SAPDXRP	-XPGSTAPXAHGVT	SAPDXRXPXG	530
Qy	630	TTT	-----	KKPAPKELAPTT	TKBPTSTSDK	PAP--TTPKGTAPTTP	KKPAPKELAPTT	677
Db	531	TAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	-----	585
Qy	678	KEPAPTTPG	-----	TAPTTLKEPAPTTP	-----	-----	KKPAPKELAPTT	723
Db	586	PPXG	STAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	644
Qy	724	KPAP--TTTK	-----	ETAPTTPKEPAPTTP	-----	-----	KKPAPTTPET	767
Db	645	RPXG	STAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	SAPDXRXPXG	-----	702
Qy	768	TTTKEPTT	-----	IHKSPDEST	PELSAETTP	KALENSPKEP	GVPTTKTPAA	813
Db	703	DXRXPXG	STAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	762
Qy	814	-TRPEWTTTAKOK	---	TTERDLRTTPE	TTTAPKMTKE	TATTTT	EKTIESKITATT	869
Db	763	DXRXPXG	STAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	SAPDXRXPXG	STAPXAHGVT	820
Qy	870	TTTQDTPPK	ITTLKTTT	LAPVTTTKTKT	ITTTTEIMNKP	PEETAPK	DKORATNSKAT	929
Db	821	APDXRXPXG	-----	STAPXAHGVT	-----	SAPDXRXPXG	STAPXAHGVT	868
Qy	930	KPTKAPKPT	STKKPTMP	RVPRKPKT	TTPTPRKMT	STMP	ELNAPT	989
Db	869	-GSTAPXAHGVT	SAPDXRP	--XPGSTAPXAHGVT	SA-PDXRPX	-----	-----	911
Qy	990	-PNSK	LVENVPK	SEDAGGAEGET	PHMLRPH	VFMEVPT	DDMDYLPRV	1048
Db	912	APXAHGVT	SAPDXRXPXG	STAPXAH	-----	-----	-----	957
Qy	1049	ETNICNGK	PVDGLTT	-LRNGTL	VAFRGHY	FWMLSFP	SPSPARKITE	1103
Db	958	VT	SAPDXRXPXG	STAPXAHGVT	SA-----	-----	-----	1003

RESULT 15  
US-08-276-967-2  
Sequence 2, Application US/08276967  
Patent No. 5851817  
GENERAL INFORMATION:  
APPLICANT: Hardy, Daniel M.  
APPLICANT: Garbers, David L. Species-Specific Egg-Binding Proteins of  
TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of  
TITLE OF INVENTION: Sperm  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, white & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/276,967  
FILING DATE: Submitted Herewith  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: UTSD:418\KIT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-787-1400  
TELEFAX: 713-789-2679

```

; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-276-967-2

```

Query Match	5.9%	Score 419.5	DB 2	Length 2476
Best Local Similarity	32.5%	Pred. No. 1.7e+18		
Matches 163	Conservative 43	Mismatches 177	Indels 119	Gaps
Qy	512	PKBPATTPKETAPTTPKKLTPPTPEKLAFTTPEKPAFTTPEELAPTTPEELAPTTPEEPTPTTPEEP	571	
Db	313	PSETSVSTERKPVAPTE-----KPTVPSIYTIPIEKPVMVHMEKDIVUT--EKPTVPT-EKP	365	
Qy	572	APTTPKAAAPNTKEPA-----PTTPKEPAPTTPEKAPTTPKETAPTTPKGTAPTTLKEP	627	
Db	366	TIPTKSTVPT--KKPIVFKEPTLPPE-GPTVPAE-RPTTPSGPAVPPKG--PTVLTE-	418	
Qy	628	APTTPKKPAKELAPTTTKETSTTSOKPAPTTPKGTAPT--TKPEPAPTTPEKAPTTPK	686	
Db	419	-----WPTSHTSEKSVHTEKPIILPTGKSTIPTEKPMVPTKRT-----TTP-	458	
Qy	687	GTAPTTLEKAPTTPKPAKELAPAT--TTKGPSFTT---SDKPAFTTPKETAPTTPEKPA	742	
Db	459	-TERTTIPAEKPTV---PIEKPMVPTERTTIPTERTIPIEKTPTVPEKLTUVPT--EKPI	512	
Qy	743	PTTEKPAFTTPTPPP-----TTSEVSTPT-----TTKEPTTIHKSPDESTPELSAET	792	
Db	513	VPTKEPIVPTPEKHPIPEKLTVLTERTTTERTTIPTKEPTVPTKEPSVPT-EKTVPT	571	
Qy	793	PKALENGKPEGVFTTKTPAAIKPEWTTTAKDKTERDLRTTETTTTAAPKMTKETATTT	852	
Db	572	-----EETIPIEKLTVPT--ERTTTTKRTTITTTIPTI-----RTTPT	613	
Qy	853	EKTTESKITATTQVTSITQDTPFKTITLTKTTLAPKVTTHKKTITTTTEINMKPEETA	912	
Db	614	ERTT-----TPTIRTTTPTERTT---IPKPKTT-----VPEKTIIPT-----ERTI	652	
Qy	913	KPKDRATNSKATTPKPOKPTKAPKKPTSTKPKTTPMRVRKPTTPPKMTSTMPELNPT	972	
Db	653	AP-----TTQP-SPLVPTQPAVWPSI-----SATTVPRTTIASCP-----PN	693	
Qy	973	SRIAEAMLQTTTRNQTNSKL	994	
Db	694	AHFERCACPVSCO-SPTPNCEL	714	

Search completed: April 26, 2002, 16:26:41  
Job time: 541 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:18:41 ; Search time 82.85 Seconds  
(without alignments)  
890.006 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1167

Perfect score: 5155

Sequence: 1 VKDNKNRTRKKKPTPKPPV.....GKPDVGLTLRLNGTLVAFRG 968

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*

1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062.5	20.6	3020	A43932	mucin 2 precursor,
2	942	18.3	1664	T18262	S-layer protein -
3	853	16.5	1489	T11108	cyst germination s
4	802	15.6	1274	T16251	hypothetical prote
5	784	15.2	1188	S49915	extensin-like prot
6	776	15.1	2187	T30826	nascent polypeptid
7	763	14.8	1367	S48478	glucan 1,4-alpha-g
8	671	13.0	1151	T18535	high molecular mas
9	667	12.9	1344	A35175	mucin 1 precursor,
10	659.5	12.8	1229	T25697	hypothetical prote
11	633	12.3	489	T11622	extensin class 1 p
12	632.5	12.3	7962	T13846	elastic titin - hu
13	632	12.3	761	C84672	hypothetical prote
14	629	12.2	3507	T34513	hypothetical prote
15	626.5	12.2	990	T15168	nucleolar phosphop
16	625	12.1	6642	T29757	protein UNC-89 - C
17	624.5	12.1	839	T75518	hypothetical prote
18	620.5	12.0	971	T19431	hypothetical prote
19	617	12.0	3570	T45025	mucin MUC5B, trach
20	607.5	11.8	801	T29018	hypothetical prote
21	607.5	11.8	924	T27923	gene LF3 protein -
22	605	11.7	379	S50125	larval glue protei
23	559.5	10.9	350	S22456	hydroxyproline-ric
24	556.5	10.8	856	T16543	hypothetical prote
25	556.5	10.8	875	S23760	polyphenolic adhes
26	555.5	10.8	1630	T25377	ascites sialoglyco
27	551	10.7	620	S06733	hydroxyproline-ric
28	549	10.6	2232	T24434	hypothetical prote
29	545	10.6	873	A47283	calphotin - fruit

30 542 10.5 369 2 S20500 hydroxyproline-ric  
31 538.5 10.4 416 2 J00465 extensin precursor  
32 533 10.3 1087 1 QFMSH neurofilament trip  
33 532 10.3 756 2 T27642 hypothetical prote  
34 530.5 10.3 1162 2 JH0557 exo-alpha-sialidas  
35 530 10.3 865 2 A47282 calcium-binding pr  
36 522 10.1 328 2 JQ0985 hydroxyproline-ric  
37 518 10.0 813 2 S70795 vsaA protein precu  
38 518 10.0 866 2 T45462 membrane glycoprot  
39 518 10.0 1072 1 A37221 neurofilament trip  
40 517.5 10.0 1459 2 T32271 hypothetical prote  
41 512.5 9.9 867 2 T45463 membrane glycoprot  
42 509.5 9.9 662 2 A45155 mucin FIM-C.1 - Af  
43 505 9.8 700 2 A54641 interspersed repea  
44 504.5 9.8 606 2 A43427 neurofilament trip  
45 504 9.8 1832 2 T31113 mucin-like glycopr

## ALIGNMENTS

RESULT 1  
A43932  
mucin 2 precursor, intestinal - human (fragments)  
N:Alternate names: mucin SMUC-41  
C:Species: Homo sapiens (man)  
C>Date: 10-Mar-1993 #sequence\_revision 12-Apr-1996 #text\_change 05-Nov-1999  
C:Accession: A49963; A45106; B45106; B33532; A61257; PQ0328; PQ0329  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t  
A:Reference number: A49963; MUID:94132002  
A:Accession: A49963  
A:Molecule type: mRNA  
A:Residues: 1-639 <GU1>  
A:Cross-references: GB:L21998  
R:Toribara, N.W.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
A:Reference number: A45106; MUID:93016075  
A:Accession: A45106  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 626-1895 <GU2>  
A:Cross-references: GB:M84131; MID:g186395; PIDN:AAA59163.1; PID:g186396  
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A:Accession: B45106  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 2037-3020 <GU3>  
A:Cross-references: GB:M84132; MID:g186397; PIDN:AAA59164.1; PID:g186398  
A:Experimental source: colon  
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,  
J. Clin. Invest. 88, 1005-1013, 1991  
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym  
A:Reference number: A43932; MUID:91358717  
A:Accession: A43932  
A:Molecule type: DNA  
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; MID:g188863; PIDN:AAA59875.1; PID:g188864  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)  
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl  
A:Reference number: A33532; MUID:89197956  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GU4>  
A:Cross-references: GB:M22405; MID:g188873; PIDN:AAA36334.1; PID:g188874  
A:Experimental source: intestine  
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

[illegible]

Qy	513	-----	512
Db	1820	NISCRATMPDVPIGOLGTVVCDVSVGLICKNEDQKGGVIMAFCLNVEINVOCCEV	1879
Qy	513	TAPTTLKEPAPTPKKPAKELAPTT--TKEPTSTSDKP---APT--TPKGTAPTTPKE	565
Db	1880	TOPTTM---TTTTENPTTPITTTTTPPTPTPTSTQSPNGLOAPPTPLISITTTTVP	1936
Qy	566	PAPTPKEPAPTTPKGTAPTTLKEPAPTPKKPAKELAPTTTSGPTST--TSDKPAPTTP	624
Db	1937	PTPTGTQTP--TTPITTTTTPPTPTPTGTQPTPVLTITTTTTPPTPTSTKSTTVTP	1995
Qy	625	KETAPTPKEPAPTTPKKP-----APITPPTPTTSEVSTPTTK-----EPTTIHK	672
Db	1996	ITTTTTPVATPTPTGTQTPMIDISTTTTTPPTPTTGTGPPPHSTAPIAELTNSP	2055
Qy	673	SPOESTPELSAEPTRKALENSKPEGVPTTKPAATKPEMTTTAKDKTTERDLRTTPT	732
Db	2056	PPESSTPQTSRSTSSPLTESTILLSLPPIAEMTSTAPSTPTAPTTTSGGHTLSPPE	2115
Qy	733	TAAPKMTKETAATTEKTSKITAATTQVTSITTTQDTPFKLTT---LKTITLAPKVTT	789
Db	2116	TTSPPGPTRGTTT--GSSSAPTPTVQTITTTSAWPTPTPLSTPSIIRTTGLRPYSSV	2173
Qy	790	KKTIITTEIMNKPTEE-----KSEDAGG-----TAKPKDRATNSKATTP	820
Db	2174	LICVLNDTYAAGEEVYNGTYGDCYFYNCISLCTLFYNWSCPSTPTPTPSK--STP	2232
Qy	821	KPKQPTKAPKKPTSTKKPTMPKVRKPKTTPTPRKMTSTMPLENTSIAEAMLOTTTRP	880
Db	2233	TPSKPSSTPKPTPGTKPPECDFDPPR-----QENETWLCDFCM--ATCKY	2278
Qy	881	NOTNSKLIVEVNP-----KSEDAGG-----AEGTPEHMLL-----	910
Db	2279	NNTVEIVKSECEPPMPTCSNGLOPVRVEDPGCCWHNECDCYCTGWDG--PHYTFDGLY	2337
Qy	911	-----RPHVFMPEVTPDMDYLPVRVNOGIIN	937
Db	2338	YSYOGNCTYVLVEIISPSVD-----NFGVIID	2364
RESULT	2		
	Tl8262		
	C:Species	Clostridium thermocellum	
	C:Date	15-Oct-1999	#sequence_revision 15-Oct-1999
	C:Accession	Tl8262	
	R:Fujino, T.; Beguin, P.; Aubert, J.P.		
	J. Bacteriol.	175, 1891-1899, 1993	
	A:Title	Organization of a Clostridium thermocellum gene cluster encoding t	
	e.		
	A:Reference number	Z18847; MUID:93209931	
	A:Accession	Tl8262	
	A>Status	preliminary; translated from GB/EMBL/DBJ	
	A:Molecule type	DNA	
	A:Residues	1-1664	<FUJ>
	A:Cross-references	EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAAA4784.1	

	Query Match	18.3%;	Score 942;	DB 2;	Length 1664;	
	Best Local Similarity	33.0%;	Pred. No. 3.2e-33;			
	Matches 276;	Conservative 83;	Mismatches 312;	Indels 166;	Gaps 43;	
y	128 VLAKPTP-KAETTKGPALTTPKEPTTTPKPEASTTPKEPTTTIKSAPTTKPAPATT	186				
b	758 VLIQAPIKAASDEPIPTDPSDETPS-----DETPS---DEPTSPDS	804				
y	187 TKSAPTTKEPATTTKEPATTTKEPA-----PTTTKEPATTTKSATTTKEPATTTPK	242				
b	805 EPAPTSEEPPIPTDPSEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSEEPPIPTDP	864				
y	243 KPAPTTKPAPTTKEPTTTPKPEATTTKEPATTTTPKEAPAPKKPAPTTTKEPAP	301				



Db 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--SDEPTPSDEPTPSDEPTPSDEPTPS 919  
QY 302 TTPKSPA--PTTKPSPST-TPKEAPATTKS-APTITKEPAPTTTKAPATTPKPSPST 357  
Db 920 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 968  
QY 358 TKEPAPTTKEPAPTTKPKAPTTKKEPAPT--TPKEAPATTTKKAPAPKAPATTTPKE 416  
Db 969 SDEPTPS--DEPTPS--DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD--- 1019  
QY 417 TAPTPKLLTPTPPKLAPTTPEKAPATTPEELAPTTPEEPPT--TPKEAPATTPKAAAP 475  
Db 1020 -----EPTPSDE--PTPSDEPTPSD---EPTPSDEPTPSDEPTPSDEPTPSDEP 1064  
QY 476 NTPKEPAPTTKEPA--PTTKPAPAPT--TPKETATTTKGTAPTTLKAPAPTTPKKAPK 532  
Db 1065 TPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS--DEPTPS 1122  
QY 533 ELAPTTTKAPT--STTSKAPAPTTPKGTAPTTKAPAPTTPKAPAPTTPKGTAPTTLKAPA 591  
Db 1123 D-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS--DEPTPS--DEPTPSDEPT 1175  
QY 592 PT-TPKKAPKELAPTTTKGTSTTSKAPAPT--TPKETAPTTKAPAPT--TPKKAPAPT 646  
Db 1176 FSETPEEP--IPTDPSDEPTPSDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPSDEPT 1229  
QY 647 TPETPPPTTSEVSTPTTKETTTIKHSDESPPELSAERTPKALENSKEPCGVPTTKPA 706  
Db 1230 DTPSDEPTPSD--EPTPSDEPT--PSDEPTPS--SDEPTPS--SETPEEP--IPTDTPSD 1277  
QY 707 ATKPEMTTAKDKTTERDLRTTPTTTAAPKMTKETATTTTEKTTESKITATTTQVTSITT 766  
Db 1278 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1337  
QY 767 QDTPFKITLLTKTTLAKVTTTKITTTTEIMNKPETAKPKORATNSKATPKQ--KP 825  
Db 1338 DEPTPSDEPT-----PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1372  
QY 826 TKAPKPTSTKPKTMVRVKPTTTPKTKMTS-----TMPELNPTSRIA 870  
Db 1373 SETPEPTPTT-----TPPSITPTSGSGSGSGSGSGGGGGGTVTSPPTPTS--- 1424  
QY 871 EAMLOTTTRPNQTPNSKLVENPKSEAGGAEPTPHMLLRPHVPMVEPTDMDYLP 927  
Db 1425 -----KPTSTAPTEIE-EPTPSDVGAIIGGHRAYLRCY-----PDGSRFP 1465  
RESULT 3  
T31108  
cyst germination specific acidic repeat protein precursor - Phytophthora infestans  
C:Species: Phytophthora infestans (potato late blight agent)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T31108  
R:Goernhardt, B.  
submitted to the EMBL Data Library, April 1998  
A:Reference number: Z20986  
A:Accession: T31108  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1489 <GOE>  
A:Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC72308.1  
C:Genetics:  
A:Gene: car-90

Query Match 16.5%; Score 853; DB 2; Length 1489;  
Best Local Similarity 32.3%; Pred. No. 1.7e-29;  
Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;  
QY 9 TKKKPTPKPPVVDEAGSLDNGDKVITPDTSTT-----QHNKVTSPKTTTAKPINR 62  
Db 338 TPYAPTEKPYDVEETTYVTEESTYAPTAKSETNAPTRMHYAHIEKPCDTEVTMVAPEET 397

QY 63 PSLP-----PNSDT-----SKETSLVKNKETTIVETKETTT---TNKQSTSD 100  
Db 398 TVAPTEETTYAPTEETTYAPTEETTYAPTEETTYPTTEETTYAPTEETTYAPTEETTYAP 457  
QY 101 GKEKTSIAKETOSIKTSIAKOLAPTSKVLAKPTPKAETTTKGPALTTPKEPPTTPPKPA 160  
Db 458 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 514  
QY 161 STTPKEPTPTTTIKSAPTTKAPAPT--TTKSAPTTKAPAPT--TTKEPAPTTPKAPAPT 217  
Db 515 EETPYEPTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 573  
QY 218 TKEPAPTTTKSAPT-----TPKEPAPTTPKAPAPTTPKEPAPTTPKAPTTPKE 267  
Db 574 TVAPTEET--YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 632  
QY 268 P-----AP-----TTKEPAPTTKAP-----APTAPKKPAPTTPKEPAPTTP 304  
Db 633 PTEETTYASTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEE 592  
QY 305 KEAPAPT--TTKEPSPPTTKAP-----APT--TTKSAPT-----TTKBP-----A 339  
Db 693 TTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 752  
QY 340 PT--TTKSAPT--TPKEPSPPTTKAP-----APTTPKEPAPTTPKAPAPTTPKEPAPT-- 388  
Db 753 PTEATYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEE 812  
QY 389 TPKEPAPTTTKKAPAPTTPKEPAPTTPKAPTTPKAPTTTPKAPTTTPKEPAPTTPKE 448  
Db 813 TPYEETTEETTYAPTEETTYAPTEET--TYPTTEETTYAPTEETTYAPTEETTYAPTEET 870  
QY 449 LAPT--TPKEPPTTPTEP-----APT-----TPKAAAPNTPKAPAPTTPKBP- 489  
Db 871 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAST 930  
QY 490 -----APTTPKEPAPTTPKAPTTPKGTAPT--TLKEPAPTTPKKAPK-ELAPT--TT 539  
Db 931 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 990  
QY 540 KEPTSTTSKAPAPTTPKGTAPTTPKAPAPTTPKAP-----APTTPKGTAPT----- 585  
Db 991 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAST 1050  
QY 586 -----TLKEPAPTTPKKAP-----APKE-----LAPT--TTKGPSTTSOKPAPTTP 624  
Db 1051 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1110  
QY 625 KETAPTTPKEPAPTTPKKAPAPTTPETPPPTTSEVSTPTTKAPT--TIHKSPDESTPELS 682  
Db 1111 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1169  
QY 683 AEPTPKA--LENSPKBP-----VPTTKTPAATKPEMTTTAKDKTTERDLRTTP-----ETT 732  
Db 1170 TEETPYAPTEETTYPTGCTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1229  
QY 733 TAAPKMTKETATTTTEKTTESKITATTTQVTSITQDTH-----PFKITTT---LKT 779  
Db 1230 TYAP-----TEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEE 1284  
QY 780 TTLAPKVTITTKTITTTIEMKNKPEETAKPKORAT--NSKATTTPKP--OKPTAPKAPT--- 833  
Db 1285 TYAPTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1344  
QY 834 -----STKKP--KTMPRVRKPKTTTPPKMTSTMPELAPTSRIAEALQTTTRP--N 881  
Db 1345 PAEESTSVSTKPCNTEETDEPTDEPTDE--PSDEPTDEPTDEPTDLPTDEPTPCDN 1402  
QY 882 QTPNSKLVENPKSDEGAC 899  
Db 1403 OGINGIGVENKVRVNNAG 1420

RESULT 4

T16251  
hypothetical protein F35A5.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
C:Accession: T16251  
R:Leimbach, D.  
submitted to the EMBL Data Library, January 1996  
A:Description: The sequence of C. elegans cosmid F35A5.  
A:Reference number: Z18485  
A:Accession: T16251  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1274 <LEI>  
A:Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GN00  
A:Experimental source: strain Bristol N2; clone F35A5  
C:Genetics:  
A:Gene: CESP:F35A5.1  
A:Map position: X  
A:Introns: 1272/2

Query Match 15.6%; Score 802; DB 2; Length 1274;  
Best Local Similarity 28.0%; Pred. No. 2.2e-27;  
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

QY 6 KNRTRK-KP-----TP-----KPPVVDGAGSLDNGD----FKVTTPDTSTTHKNKSVTS 50  
DB 274 KNTKAKKPPWEDETVEEVKPPVEKKAPVLKKDPAPAKARDPSPSKAAPKVPES 333  
QY 51 PKITTAAPKI-NRPSPSPNSDTSKETSIVNKETTIVETKETTITTTNKQSTGDKKETSATK 109  
DB 334 SPVVPPTPVKNPVKKYKPPWVDDEPAEVPKPSAEPKKTPLVKRKEPSSSTTPSSDPS 393  
QY 110 EQTSIEKTSAKDLAPTSKVLAKPTPAE-----TTTKGPA-----LITTP 148  
DB 394 PKKAAPAVKPRSSPKKATPLQADPKRAQVPPTPVKNPVKKYKPPWVDDEDPVEEVKQP 453  
QY 149 KEPTPTTP-----KEPASTTPKEP-----TP-TTKSAPTTKPEPAPTTTKSAPTTKPEPAP 199  
DB 454 EAPAKTTPVLKKEPAAKDTAKATSKTPTETPEKDPVKPRSSPKKVAAKPDSAGAPA- 512  
QY 200 TTTKEPA-----PTTPKEPAPTTTKEPAP-----TTKSAPTTTP 233  
DB 513 TPVKNPVKWRRPPWEDDETPADDSKPTDAKKTSLAKDPAPAKESLKPADTKAPAKP 572  
QY 234 KEP-----APTPKKAPATTPKE-----PAPTPKEPTPTTPKEPAPTTTKEPAPTTTPK-EP 283  
DB 573 RDPSPKKVAPTAPEKKTTPVLAKKEPAGADSKTKEPEKSKRDPSPKKAVPAKVPKTEV 632  
QY 284 APAPKPPAP-----TTPEPAPTTTKEPAPTTTKEP-----SPTTPKEPAPTTTKSAPT 333  
DB 633 APAAVKKPEPISKPDATPKAEPNSPVVP-PTPVKNPVKKWKPWEDDDAPAKVSLPE 691  
QY 334 TTKEPAPTTTKSAPTTTP-KEPSTTTTKEPAPTTTPK-----EPAPTTTPKAPAP-----TTPEPAPT 388  
DB 692 PEKK--TPVLAKKAPTDPSEAAADPVSGSSKDPKLAKEAPKPRDPSPMKAVPIKPAKP 750  
QY 389 TPKEPAPTTTTPKAPTA--PKEPAPTPKETATPTTKLPTTP-----EKLAPTTTPKPA 442  
DB 751 T--EVPAVVKKEPPEKAVKRDPSKKAK--AEPNSP--VVPPTPVKNPVKKWKPWEDDDA 805  
QY 443 PTTPEELAPTTPEP-----TPTTPEPAPTTTPKAAANTPKPEPAPTTTKEPAPTTTPK-EP 496  
DB 806 PAEPVNVPEKKTTPVLAKKTPVKPRDPSKKAVPAKPSKTDAPPVSVKKKEPVPKPE 865  
QY 497 PAPTTPKETAPTTPKGTATTTTKEPAPTTTPKPAKELAPT--TTKEPTSTSDKPAPTTP 555  
DB 866 PSPKKAEPNSPVVP-----PTPVKNPVKKW--KPPWEDDEPTEEVKKPSE--PEKKTPLVA 918  
QY 556 KGATPTTPKEPAPTTTKEPAPTTTPKGTATTTTKEPAPTT-----TPKKPAP-----KEL 603  
DB 919 K-KEPEKPD-APKVAAPKRDPSKKAVPE--KEPAKVAAPKRDLSFKKAIPANTQEA 974

QY 604 APTTTKGTST-----TTSDKP-----APTTPKET-----APTTPKETAPTTPKKPAPT 646  
DB 975 PPTPVKNPVKKWKPWEDDDDEPAEPVSAPEKKTPLVAKKAPAKPRDP---SPKKAADV 1031  
QY 647 TPETPTPTTSEVSTPTTTTKEPTTIHKSP-----DESTPELSA-EP---TPKALENSPKPEG 698  
DB 1032 AAK-PDPKIPEV-PPTPVKNPVKKWKPWEDDDDEPSEPVSAPEKKTPLVAKKAPTRPA 1089  
QY 699 V-----PTTKTPAAT-----KPEMTTTTAKDKTTTERDLRTTPTTT--TAAPK 737  
DB 1090 TAPDSEAAADPVSGTSPKDPKLSKKAPVKKPPTTDPDKDLKPSAKKAPKAPAPAKP 1149  
QY 738 MTKETATTTTEKTESKITATTTQVSTTTTQDTPPTPKITTLTKTTTLAPKVTTTKTITTTTE 797  
DB 1150 KKKPVWDDDDPDEADFTVPAPSKKPDTEADPADPLG-----GPKTKDPK----- 1193  
QY 798 IMNKPEETAKPKDRATNKAITTPKQKPKAPKPKPTTSTKTKPKTMPRVKPK----- 848  
DB 1194 -LNKKAPAEKPTK-----PKPEVSKPEPKPTTEPPKP-AAPKKWKPWEDDDDEDE 1243  
QY 849 ---TTPTPRKMTSTMP 861  
DB 1244 ADETPAPKAPKPDTEP 1259

RESULT 5

S49915  
extensin-like protein - maize  
C:Species: Zea mays (maize)  
C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
C:Accession: S49915  
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.  
submitted to the EMBL Data Library, June 1994  
A:Description: Pex genes: pollen-specific genes with extensin-like domains.  
A:Reference number: S49915  
A:Accession: S49915  
A:Molecule type: DNA  
A:Residues: 1-1188 <RUB>  
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Query Match 15.2%; Score 784; DB 2; Length 1188;  
Best Local Similarity 29.9%; Pred. No. 1.2e-26;  
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

QY 13 PTPKPPVVDGAGSLDNGDFKVTTPDTSTTHKNKSVTSKITTAKTINRPSL----PPN 68  
DB 460 PTPHSPAD-----DYVPTTPVPGKSPSPATSPSQVQPPAASTPPPSLVKLSPPQ 510  
QY 69 SDTSKETSIVNKETTIVETKETTITTTNKQSTGDKKETSATKETSATKDLAPTSKV 128  
DB 511 APVG-----SPPPPVKTTSPAPIG-----SPSPPPPVSVV 541  
QY 129 -----IAKPTPKAETTTKGPALTTTPKEPTTPTPKEPASTTPKEPTTTTTSAPTTKEPA 183  
DB 542 SPSPPVKSPPPAPVSGSPPPPEKSPPPAPVAPVAPSPPPVKSPP---PPPTLVASPPPVKSP 599  
QY 184 PTTTKSAPTTTP--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTP--KEPAPT 240  
DB 600 PAPVASPPPVKSPPPPTPVASPPPPAPVASPPPMKSPPPPTPVSSPPPPPEKSPPPPP 659  
QY 241 PKKPAPTTTKEPAPTT-----PKE---TPPTTPKEPAPTTTKEPAPTTTKEPAPTAKPK 291  
DB 660 PAKSTPPEEYPTPTTSVKSSPPPEKSLPPTLIPSPPPQEKPTPTPSTPSKP--PSSPEK 718  
QY 292 APTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTPK 351  
DB 719 SP--PKPEVSSSPQTP-----KSPPPAPVSSPPPTPVSSPPPALAPVSSPPSVKSSP---- 768  
QY 352 EPSPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTAKPEAP 411



A:Accession: A26877  
A:Molecule type: DNA  
A:Residues: 1-242 <YAM>  
A:Cross-references: EMBL:M16164; NID:gl72522; PIDN:AAA35014.1; PID:gl72525  
A:Accession: B26877  
A:Molecule type: DNA  
A:Residues: 762-1331 <YA2>  
A:Cross-references: EMBL:M16165; NID:gl72523; PIDN:AAA35015.1; PID:gl72526  
R:Pardo J.M.; Ianez E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
FEBS Lett. 239, 179-184, 1988  
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae  
A:Reference number: S27281; MUID:89031230  
A:Accession: S27281  
A:Molecule type: DNA  
A:Residues: 1-31 <PAP>  
A:Cross-references: EMBL:X13857; NID:q4551; PIDN:CAA32069.1; PID:q4552  
R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal growth of *Candida albicans*  
A:Reference number: JG6123; MUID:96323237  
A:Accession: JG6123  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1367 <LAM>  
A:Cross-references: GB:U30626; NID:gl1304386; PIDN:AAC49609.1; PID:gl1304387  
C:Genetics:  
A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
A:Cross-references: MIPS:Y1R019c; SGD:S0001458  
A:Map position: 9R  
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
F:5-21/Domain: transmembrane #status predicted <TM1>  
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 14.8%; Score 763; DB 1; Length 1367;  
Best Local Similarity 28.5%; Pred. NO. 1.le-25;  
Matches 286; Conservative 98; Mismatches 465; Indels 154; Gaps 43;

QY 35 TPPTDSTTHQNKVSPKTTAKPKNP-----RPSLPNDSRDKTSLVNKETTVE 86  
DB 236 STSESSTTSSTSSSTSTAPATPTTCTCKEPPPTTCTCKEPPPHDHTTPC 295  
QY 87 TKETTTTNNQSTDGKKTSAKETSQIEKTSKADLAPTSKVLAKPTPKAETT--TKGPA 144  
DB 296 TKKKTTSK-TCT---KKTTTPVTPPS-SSTTESSAPV-----PTPSSSTTESSAPV 344  
QY 145 LTPKE-----PTTTPKEPASTTPKEPTTIKSA-----TTPKEPAPTTTKSAPTTP 194  
DB 345 TSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPTPSSSTTE 404  
QY 195 KEPAPTTTKAPAPTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTKEPAP 254  
DB 405 SSAPVTS---TTESSAPVTS---STTESSAPV---SSTTESSAPVTSSTTES 453  
QY 255 TTPKEPTP---TTPKEPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAP 307  
DB 454 SSAPVPTPSSSTTESSAPV---SSTTESSAPV---VPTPSSSTTESSAPVTSSTTES 509  
QY 308 APPTTKPSPTPKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTT 363  
DB 510 APVPTPS---SSTTESSAPAPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTES 568  
QY 364 ----TTPKEPAPT-TPKKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTPKETA 418  
DB 569 VTSSSTTESSAPVPTPSSSTTESSAPVPT---PSSSTTESSAPAPTPPSSSTTESSA 624  
QY 419 PTPPKLPTTPKEKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTT 472  
DB 625 PVT---SSTTESSAP-VPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTES 679  
QY 473 AAPNTPKEPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTTTKAPAPTT 531

DB 680 SAPVT---SSTTESSAPVT---SSTTESSAPVPTPSSSTTESSAPVPTPSSSTTE 731  
QY 532 KELAPTTTKEPTSTSDKAPAPTTPKGTAPPTKEPAP---TTPKEPAPT-TPKGTAAPTTL 587  
DB 732 SSSAPVTPPS-SSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTES 790  
QY 588 KEPAPTTTPKKAPKELAPTTTKGPTSTTSDKAPAPTTPKETAPTTTPKEPAP-----TTPKK 642  
DB 791 SAPVPTPSSSTTESSAPVPTPSSSSNITSSAPSTPSSSTTESSSVPTPSSSTTES 850  
QY 643 PAPTTPETPPPTTSVSPT-----TTPETTHKSPOESTPELSAETPKALENSPREP 697  
DB 851 SAPVSSSTTESSAPVPTPSSSSNITSSAPSSI---PFSSTTESFTGT-TVPSSSKYP 906  
QY 698 GVTPTTKTPAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETATTEKTTESKITAT 757  
DB 907 GSQETSVSSITETIVTKITTSVTSTTTTITTVCGSTGNSAGETSGCSPKTVIT 966  
QY 758 --TTOVSTTTQDTPPFKITTLLKLTLLAPKVT---TKKTIITT-EIMNKPEETAAPKDR 811  
DB 967 VPTTTTTSVTSSTTTITTVCGSTGNSAGETSGCSPKTIITTVPCSTSPSETA----- 1021  
QY 812 ATNSKATTPKPOKP-----TKAPKKPTSTKKKPMRVRPKPTTPPR 854  
DB 1022 ---SESTTSPPTPVTVTVTVTVEYSTKPGGEIITTEVTKNIPTYLTITTAFTP- 1077  
QY 855 KMTSTMPELNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRHV 914  
DB 1078 -SVTVTVNFTPT-ITTVVCSGT-----NSAGETSGCSPKTVITVPCST 1122  
QY 915 FMPEVTPMDYLPRVPGNOGIIINMLSDETNTCNKQPDVGLTT 957  
DB 1123 GTGEYTTTATLLVTTAVTTTVVTTTESSTGNSA-GKTTTGYTT 1164

RESULT 8  
TI8535  
high molecular mass nuclear antigen - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: TI8535  
R:Shimada, K.; Harata, M.; Mizuno, S.  
J. Cell Sci. 110, 3031-3041, 1997  
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of ch  
A:Reference number: Z18955; MUID:9803440  
A:Accession: TI8535  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1151 <SHI>  
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

Query Match 13.0%; Score 671; DB 2; Length 1151;  
Best Local Similarity 25.9%; Pred. No. 7.3e-22;  
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

QY 14 TPKP-----PVDGAGSLDNGDKVTTPTDSTTHQNKVTS--PKITTAAPINPRPS 64  
DB 113 TPPPSSGAGTTPPSOGAAGAPKGDGTAQPSGTSKADGKPAQDVPKATTA-ATEARD- 170  
QY 65 LPNDSSTKETSILVNKETTVEKTTTNNK--QTSTDGKETTSAKETQSIKTSKDL 122  
DB 171 -----ASAASPTVPKATAEATAVTAASQSAPKAATDAATAA---SOSAPKATV-EV 219  
QY 123 APTSKVLAKPTEKAEETTTKGPALITTPKEPTP-TTPKEPASTTPKEP-----TPTTIKSAPT 177  
DB 220 KPAAAVAKEAKAVTAAAPAKATAEAKPAPVSTPIPCSSAEAKPLTAASATASKA--T 277  
QY 178 TPKEPAPTT-----TKSAPTTTPKEPAPTTTK----- 203  
DB 278 AEAKVPVATASLMATKVTAEAKPAPSPVPKATTDOTKAVTATATKAGPDVKAAYCAEA 337  
QY 204 EPAPTTTPKEPAPTTTKAPAPTTTKSAPTT-----PKEPA-----PTTPKKPAPTTTPK 250

```

338 KPAPPPPPQOLKAAAAAAPTGTGLKAPATAPPHGSPRANSHTVTVTPNNVPRAAAATVP- 396
251 EPAPPTTKPEPTPT---KPEAPPTTKPEAPPTTKPEAPPTAPKPKAPPTTKPEAPPTPKPE 306
397 -TAGAVPKASTGTTAAAQOQVPV---KAPVTVPPSQQAQVPRAAATAAA-----APVTPOQ 448
307 P-----APTTTKPEPSPTTKPEAPPTTKTSAPTITTKPEAP-----TTTTKSAPTITTKPEPSPTTKE 360
449 PVTKAATTTNATPPQIPKAAATTTTATVTPQOPIPKAGTDAAPPAVPKAPSDGRAAT 508
361 P-----APTTTKPEAPPTTKPKAPPTTKPEAP-----TTTKPEAPPT-TTKTKPEAPPTAP 406
509 PGVNAATDPQKPPPTQSPVSAVTEPKQPRAAAPPSNEATPAVPSPSNLKSLPTTIP 568
407 K-----EPAPPTPK---ETAPTTKKLTPTTPEKLAPTTTPKPAPTTPEELA----- 450
569 KPVPMLALTPQPVTAQWVQLAAATKPSIVPKASPK-ALMTPPPPPPGGLPRALAAAALLG 627
451 -PTTP-----EPTPTTP-----EPAPPTTP-----KAAAPNTPKE----- 480
628 LPSSPVASAMHAKVTPRPLPASVPVMAASPASGLPDAARVALATNAASPGARPEAAGNG 687
481 ----PAPTTKPEAPPTTKPEAPPTTKPEAPT-----TPKGTAPT----- 516
688 TLMAPGAANTONAPICGAAGAQTAPMGAAHTHVPWGAGGATQMSPTGAANTHMSPIGA 747
517 -----TLKEAPPTTKPKAPKELAPTTTKEP-----TSTTSKPKAPPTTP 555
748 GGATQSPMGAAANTQSPMGATTTQSPMGAAATQSPMGAAATQVATISAGNTMQVSP 807
556 KG--TAPTTKPEAPPTTKPEAPPTTKPGTAPTTKBP--APTTPKKPAKPKELAPTTTKGP 611
808 MGAATPPQTPSVCAATTP-QPSPM---GAAITLMSPMGAATTPQ---PSPGAVTTQPP 859
612 ----TSTTSKPK-APTTPKET---APTTPKEP---APTTPKKP-APTTPETPP----- 652
860 PMAATNTTQPPMAASTPOSTPMGAATTTQSPPMGATTTQSPPMGASTPQAPPTVAGSPT 919
653 PTTSEVSTPTTKPEPTTIHKSDESTEPCL--SAEPTPKALENSPKPPGVTTKTKPAATKP 710
920 PPPPIPPSTAGTSOPMKSPPPPDPKAPSAQAATSPAAHVANASPGV-TAVSPA--p 975
711 EMTTAKKTDRLRTTETTTAAPKMT-KETATTTEKTTESKINATTTQVTTSTTTQDT 769
976 IGVTEASPSADGARLSPGTAATDGPKASPAATADVTEAATD-VTAATAVPA-----EA 1029
770 TPEKITLTTTLAPKVTTKTITTTTEIMNKPETEAKPKDRATNSKATTPKQPKTKAP 829
1030 AP-----TKAKRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSNPASPA--- 1077
830 KKPTSTKKPKTWPVRVKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLV 889
1078 -----PVGDDGQQQMTPGNAQSVPP-----VTEAAVQ----- 1104
890 EVNPKSEDAGGAEGE 904
1105 EAAAAAAGAAAGAE 1119
```

## RESULT 9

A:Accession: A35175  
A:Molecule type: mRNA  
A:Residues: 1-952,1033-1344 <LIG1>  
A:Cross-references: GB:J05288; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124;  
A:Experimental source: splice form A  
A:Note: Genbank entries HUMEPISIAL and HUMEPISIA2 present only the amino-and carboxyl  
A:Accession: B35175  
A:Molecule type: mRNA  
A:Residues: 1-19,29-952,1033-1344 <LIG2>  
A:Cross-references: GB:J05288; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129;  
A:Experimental source: splice form B  
A:Note: Genbank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl  
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhlig, T.; Peat, N.; Burch  
J. Biol. Chem. 265, 15286-15293, 1990  
A:Title: Molecular cloning and expression of human tumor-associated polymorphic epith  
A:Reference number: A35886; MUID:90368715  
A:Accession: A35886  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-19,29-992,1033-1344 <GEN>  
A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870  
A:Note: Genbank entry HUMUCAB includes one copy of the tandemly repeated sequence  
R:lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.  
J. Biol. Chem. 265, 15294-15299, 1990  
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.  
A:Reference number: A35887; MUID:90368716  
A:Accession: A35887  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>  
A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599  
A:Note: Genbank entry HUMPANMU contains four fewer copies of the tandemly repeated se  
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky,  
Eur. J. Biochem. 189, 463-473, 1990  
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen  
A:Reference number: S10571; MUID:90276413  
A:Accession: S10571  
A:Molecule type: mRNA  
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>  
A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054  
R:Wreschner, D.H.  
Submitted to the EMBL Data Library, March 1990  
A:Reference number: S40293  
A:Accession: S40293  
A:Molecule type: mRNA  
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <A>  
A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054  
R:Abe, M.; Siddiqui, J.; Kufe, D.  
Biochem. Biophys. Res. Commun. 165, 644-649, 1989  
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associa  
A:Reference number: A36735; MUID:90088473  
A:Accession: A36735  
A:Molecule type: mRNA  
A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>  
A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543  
R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu  
J. Biochem. 112, 609-615, 1992  
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut ag  
A:Reference number: JX0235; MUID:93123189  
A:Accession: PX0066  
A:Molecule type: mRNA  
A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037;1038-1057 <MAS>  
A:Experimental source: gastric carcinoma cell  
R:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.  
FEBS Lett. 356, 130-136, 1994  
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokin  
A:Reference number: S51026; MUID:95080414  
A:Comments: annotation  
A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region  
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 12  
partial repeats. The repeat shown is defined by SmaI nuclease sites.  
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively  
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48

C:Genetics:  
A:Gene: GDB:MUC1; PUM  
A:Cross-references: GDB:120705; OMIM:158340  
A:Map position: 1q21-1q23  
A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3  
C:Superfamily: polymorphic epithelial mucin  
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphism  
F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>  
F:1-62/Region: mucin 1 amino-terminal non-repetitive  
F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>  
F:1-19-29-32/Domain: signal sequence #link PREB #status predicted <SIGB>  
F:1-19-29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>  
F:1-19-29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form  
F:138-1017/Region: 20-residue repeats (GSTAPPAHGVSAPDTRPAP)  
F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive  
F:1245-1272/Domain: transmembrane #status predicted <TRM>  
F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 12.8%; Score 667; DB 1; Length 1344;  
Best Local Similarity 27.8%; Pred. No. 1.2e-21;  
Matches 305; Conservative 77; Mismatches 459; Indels 256; Gaps 56;

QY 77 LTVNKETTVETKETTNNK--QTSTGCKEKTTSKAKETQSIKTSKAD-LAPTSKVLAKPT 133  
DB 15 LTVLTATAPKPAVTGSGHASSTPGGEKETSATQSSVSPSTERNAYSMTSSVLSHS 74

QY 134 P-KAETTTKGP--ALTTKPEP-----TPPTKPEASVTP-----KE 166  
DB 75 PGSGSSSTQGDVTLAPATEPAGSAAATGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 167 PTPTT-----IKSAPTTKPEAPTT-----KSAPTTKPEAPTTKPE-----A 206  
DB 135 PAPGSTAPPAHGVSAPDT--RPAGSTAPPAHGVSAPDT--RPAGSTAPPAHGVS 190

QY 207 PTTPKPEAPTTKPEAPTTKSAPTTKPEAPTT-----KKPAPTTKPEP----- 252  
DB 191 PDT--RPAGSTAPPAHGVT-SAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHG 247

QY 253 --APTTPKEPTPTTP-----KEPAPTTKPE-----APTTPKPEAPTPAK----- 289  
DB 248 TSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHG 307

QY 290 KPAPTTKPEAPTPK-----EPAPTTKPEPTTKPEAPTT--TKSAPTTKPEAPTT 342  
DB 308 TSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHG 367

QY 343 TKSAPTTKPEPTTKPEP-----APTTPKPEAPTPK-----KPAPTTKPEAPTPK- 391  
DB 368 T-SAPDT--RPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPA 424

QY 392 ---EPAPTTTKKAPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 447  
DB 425 HGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPA 484

QY 448 ---ELAPTTPEETPTTP-----EPAP--TTPKA-----AAPNTKPEAPTPK- 487  
DB 485 HGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPA 544

QY 488 ---EPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 534  
DB 545 HGVSAPDTRPAGSTAP-----PAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAGST 600

QY 535 APTTTKEPTSTSDKAP--TTPKGAAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 588  
DB 601 AP-PAHGVSAPDTRPAGSTAPPAHGVS-----APDTRPAGSTAPPAHGVSAPDTR 655

QY 589 EPAPTT-----KKAPKELAPTTKGTSTSTSDKAP--TTPK-----ETAPTT 632  
DB 656 APGSTAPPAHGVSAPDTRPAGSTAP-PAHGVSAPDTRPAGSTAPPAHGVSAPDTR 714

QY 633 KEPAPTT-----KKAPTTPTPTPTTSEVSTPTTKPTT-----IHKSPD 675

DB 715 PAPGSTAPPAHGVSAPDTRPAGS--TAPPAHGVSAPDTRPAGSTAPPAHGVSAPD 772  
QY 676 ESTPELSAETPKALENSKPEGVPTTKTPAA-----TKPEMTTKAKD---TTERD 724  
DB 773 TRPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPD 832  
QY 725 LRTTPETT-----TAAPKMKETATTTKTESKITATTTQVSTTTTQDTP--FKIT 775  
DB 833 TRPAGSTAPPAHGVSAP-----DTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVT 888  
QY 776 TLKTTTTLAPKVTTTKTITTT--EIMNKPEETAKPKDRATNSKATTPKPKPTKAPKPT 833  
DB 889 SAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAP--GSTAPPAHG 946  
QY 834 STKKPKTMRVRKPKTTPPKMTSTMPENP-----TSRIAEAMLOT--TRP---NQTN 885  
DB 947 VTSAPDTRP--APGSTAPPAHGVSAPDTRPAGSTAPPAHGVSAPDTRPAGSTAPP 1003  
QY 886 SKLVEYNPKSEDAGGAEGETHMLLRPHVFMPEVTPDMDYLLPRVNOGIIINPMLSDETN 945  
DB 1004 AHGVSAPDTRPAGSTAPPAH-----GVTSAPDNRPALGSTA---PPVHNVTIS 1049  
QY 946 ICNGKPVGGLTTLRNGT 962  
DB 1050 ASGSASGSASTLVHNGT 1066

RESULT 10  
T25697  
hypothetical protein F16F9.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T25697  
R:Fullon, B.  
submitted to the EMBL Data Library, August 1996  
A:Description: The sequence of C. elegans cosmid F16F9.  
A:Reference number: Z20071  
A:Accession: T25697  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1229 <FUL>  
A:Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9.2  
A:Experimental source: strain Bristol N2; clone F16F9  
C:Genetics:  
A:Gene: CESP:F16F9.2  
A:Map position: X  
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 12.8%; Score 659.5; DB 2; Length 1229;  
Best Local Similarity 29.0%; Pred. No. 2.4e-21;  
Matches 307; Conservative 78; Mismatches 339; Indels 333; Gaps 52;

QY 33 KVTT-PDTSTTQ--HNKVSTSPKIITA-KPINRPSLPNSDT-----SKETSITVN 80  
DB 119 KVTTSDASTNAPTCKDSTTPEITIGIVINSKSESVDMTSTTRFSTLSPTTELLS 178

QY 81 KETTVEKETTNNKOTSTDGKEKTTSAKETSIEKTSKADLAPTSKVLAKPCKAETTT 140  
DB 179 PEFVSTDSSTST-EQTSNDNTEIASPMETNTTTEATTSVEPSVSTLAS---EDETIV 234

QY 141 KGPALTPKEPTTPKPEASTTPKPEPTTKSAPTTKPEAPTTTKSAPTTKPEAPT 200  
DB 235 TATAEST-----TTVIAEVSTTTTEEPPTT-----AESTTKKSTT---KAPA 272

QY 201 TTKEAPATTPKPEAPTTKPEAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 256  
DB 273 TTEEPPTTTEE--VTTTAEATSTTTSSETSTEBK---PTTPLIDNKIAGPATGK---PET 324

QY 257 PKEP--TPTTPKPEAPTTKPEAPTTKPEAPTAAPKAPPAAPKAPPAAPTTKPEAPTTKE- 314  
DB 325 THFPVGTGTFN--FDATATPTFFVAKSEDKMTLSKTAATETTQQTTEVT-DGPEKETKNV 381

Qy	315	----	PSDTPPKPEPAPTTT	KSAP-----TTTK	EAPT-----TTKS	APTPTTKPEPSPTTK	359	
Db	382	SI	PIPTVPLVETTS	TTSTASKESDGF	TTLLKVVTTADSD	STESATTVKPFNEETTK	441	
Qy	360	E--	PAPT-----TPK	-----EPAP	TPPKPAP-----	-----TPPK	383	
Db	442	SHV	VPKPTKGT	VKVTPKLEL	SFDEPTEIT-KAPHPG	GLLEKKTVHFLV	SDFNARYSEAK	500
Qy	384	E-----	PAPTTPKEPAP-----	TTTKKPAPT	A-----PK	EPAPT-----PK	STAP	419
Db	501	ENDD	NHLDYNHV	REAKEPTTEES	TSEBVTTEEPANT	GNPPTENPTTEO	PTSTAE	560
Qy	420	TPPK	KLTPPTPEKLA-----	PTTPEK	PAPTPPELAPTPEEP	TPPTPEEPAPTTP	KAAAPN	470
Db	561	STTAL	PFTTEQVNTVEEPT	TAEKSTATQ-----	KPTTQESVST---	EKTSTTKKA---	S	610
Qy	477	TPKE	PAPTTPKEPAPTTPKE	PAPTTPKETAPTTPK	GTAPTTLKEPAPTTPK	PAPTPPKPAPKELAP	536	
Db	611	TTEE-	-PTTDEPTTTT-----	ESSTGKAPTPEL	STTSBETTTTELKI	ITTE-----GS	657	
Qy	537	TTTK	KEPTSTTSOKPAP-----	TTTPKG	TAPTTPKEPAPTTPK	PAPTTPKGAPTTLKEPAP	592	
Db	658	TTTEE	PTTAIFAESA	GLIITDEETSTST	PTPEITSKE-----	IVTESAIQT	SVSVVES	715
Qy	593	TTP-----	-----	KPAPKELAPT	TTKGP-----TS	613		
Db	716	STP	QLPERMKA	IVNFKHINLEVL	EKKRLLKEKEST	TGSDSETTTVVAEN	IDEVTT	775
Qy	614	TTSD	KAPTTPKETAPTTPKE	PAPTTPKPA	PTTPE-----TPPT	TSE-VSTPTTKET	TTI	670
Db	776	TEKE	VVQTPPTTEKSTTQ	EBTTTTTTTTTEKT	TSKTTEKPTTSES	ATTETTTSEPT-	834	
Qy	671	HKS	PDESPELS	AEPTPKALENS	KEGVPPTTKT	PAATKPEMTTAKD	KTTERDLRTTPE	730
Db	835	----	TEST-----	-----TVDT	SATTESSAAETTT	ISAE---TSE	865	
Qy	731	TTTA-----	-----	APKMTKE	PATTTEKTTE	751		
Db	866	TTTT	SESAFITG	ESPENTALQ	SSQSQBENESSA	EKPGARRDFVPK	HKHTVKPAETTS	925
Qy	752	SKT	ATTTO-VTST---	TTQD	TPKINTLKTTL	LAPKVT---TTK	KTITTEIMNKP	804
Db	926	VAAS	TTTTPTTTEKSTTLE	PTPEATLNEV	TGAPVTPGAP	VDETTINT	LELSK---	982
Qy	805	TAKP	KDRATNS	KATTPKQO-----	-----KPT	KAKPTSTKK	PKTPRVRKP	847
Db	983	-----	INNTQISQ	PKPTDISKTAL	SLSLGSLG	FTKAPMPTI-----	1022	
Qy	848	KTT	PTPRKWT	STMPELN-----	-----PT	SRIAEA	872	
Db	1023	HTTT	DAAFVATEAS	LNQSDSKK	IIIDEA	OPTDEIRRA	1059	

RESULT 11  
T11622

extensin class 1 precursor - cowpea  
 C:Species: *Vigna unguiculata* (cowpea)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_date 17-Nov-2000  
 C:Accession: T11622; S54155  
 R:Arсенијевић-Maksimovic, I.; Broughton, W.J.; Krause, A.  
 Mol. Plant Microbe Interact. 10, 95-101, 1997.  
 A:title: Rhizobia modulate root-hair-specific expression of extensin genes.  
 A:Reference number: Z17301; MUID:97153574  
 A:Accession: T11622  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-489 <ARG>  
 A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015936  
 A:Experimental source: sub\_species Red calcoona  
 R:Arсенијевић-Maksimovic, I.; Broughton, W.J.; Krause, A.  
 submitted to the EMBL Data Library, April 1995  
 A:Description: A class of root-hair specific extensins involved in rhizobium

A:Reference number: S54155  
A:Accession: S54155  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 326-489 <AR2>  
A:Cross-references: EMBL:X86030; NID:g791149; PID:g791150  
C:Genetics:  
C:Gene: Ext26G  
C:Superfamily: hydroxyproline-rich glycoprotein  
C:Keywords: glycoprotein; hydroxyproline  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-489/Product: extensin class 1 #status predicted <MAT>

Query Match 12.3%; Score 633; DB 2; Length 489;  
Best Local Similarity 31.2%; Pred. No. 1.4e-20;  
Matches 149; Conservative 33; Mismatches 252; Indels

[illegible]

RESULT 12  
I38346

elastic titin - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
 C:Accession: I38346  
 R:Labeit, S.; Kolmerer, B.  
 Science 270, 293-296, 1995  
 A:title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.  
 A:Reference number: A57430; MUID:96026330  
 A:Accession: I38346  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-7962 <RES>  
 A:Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427  
 C:Genetics:  
 A:Gene: GDB:TTN  
 A:Cross-references: GDB:I27867; OMIM:188840  
 A:Map position: 2q31-2q31

Query Match	12.3%	Score 632.5;	DB 2;	Length 7962;
-------------	-------	--------------	-------	--------------

Query Match 12.3%; Score 632.5; DB 2; Length 7962;





QY 503 KETAPTPKGTATTTLKEPAPT--TPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAP 560  
Db 481 TSPPPVQP-----PPVQKPPPTTSPVVKPPPIQKPP-----PTSPPIKPPVVKPPPTPT 532  
QY 561 TTPKEPAPTTPKEPAPT--TPKGTATTTLKEPAPT--TPKKPAPKELAPTTTKEPTSTTSD 617  
Db 533 SPPIK-PPVHKPPPTTSPPIKPPPIHKKPPPTTSPPIKPPVVKPPPTTSPPIKPP 591  
QY 618 KPAPTTPKETAPTTPKEPAPTTPKKEPAPT--TPETPPPTTSEVSTPTTK--EPTTIHKSP 674  
Db 592 VHKPPPTTSPPIK-PPVHKPPPTTSPPIKPPVVKPPPTTSPPIKPPVVKPP 647  
QY 675 DESTPELSAEPKPALENGKPEGVPTTKTPAAKPEMTTAKDKTERDLRTPTETTTA 734  
Db 648 ---TPTTSPPIKPPVQKPPPTTSPVVKPPVQLPP--TPTTSPVVKPPVQVPPPTTYS 703  
QY 735 APKMTKETATTTTEKTESKITATTTQVSTTTQDTP 771  
Db 704 PPVKKPPVQVPPPTTSPPIKPPVQVPPPTTTPSP 740  
RESULT 14  
T34513  
hypothetical protein ZK783.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34513  
R:Favella, A.; Vaudin, M.  
submitted to The EMBL Data Library, August 1994  
A:Description: The sequence of C. elegans cosmid ZK783.  
A:Reference number: Z21536  
A:Accession: T34513  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-3507 <FAV>  
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1  
A:Experimental source: strain Bristol N2; clone ZK783  
C:Genetics:  
A:Gene: CESP:ZK783.1  
A:Map position: 3  
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;  
3504/1  
Query Match 12.2%; Score 629; DB 2: Length 3507;  
Best Local Similarity 24.1%; Pred. No. 1.2e-19;  
Matches 252; Conservative 158; Mismatches 416; Indels 218; Gaps 45;  
QY 10 KKKPTKPPVDEAGGLDN-GDFKVT-----PDTSTQHNVKYSTSPKITTA-----K 57  
Db 1942 KQPNREKIEIDENSSNSGQEKPTTKGIVSSTSATSSSTAEAPHVTTTSSSTTSNK 2001  
QY 58 PINPRESLPN---SDTSKETSIVNKEIV-ETKETTITNQTSDGKEKTSIAKETQS 113  
Db 2002 DMTSSKS-PENVMTSESPEVSTSSKSTAGETTVSTSPSSSEAPLTSSPATITEV 2060  
QY 114 IKTSAKDIAP---TSKVLAKTPKAETTTKGPAITTTKEPTTTPKEPASTTPKEPPT 169  
Db 2061 ITSSVKSSTTPREESSEITVLKSSKSPVETSSVKSSTSPS-TTSQSVTSTVPTSKS 2119  
QY 170 TTIKS-APTTPKEPATTTKSAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTKS 228  
Db 2120 TVLSEAPVTSPTSEVHT--SSETPKPSLSASSTGDTNSTTPTSSSLASVKSTSAPEGS 2178  
QY 229 APTTPKEPATTTPKPAITPK-EPAPTTPKEPTTPKEPATTTPKEPAPTTPKEPAPT 286  
Db 2179 ASVAPVKLSLSPDVQSPTKTFDATESSTVQASESTTSVKSTSEPESHVTKLSITS 2238  
QY 287 APKKPAPTTPKPAITTPKEPATTTPKPSPTTP--KEPAPTTPKSAPTTPKEPATTTK 344  
Db 2239 NPSSSVPTVSPKSTPVPE-----STEQPTSTPSCQSLTPMNSNEVLTTSEPHVLSSS 2293  
QY 345 SAP-----TTPKEPSPTTTPKEPATTTPKEPATTTPKPAITTPKPAITTPKEPAPTTP 397

Db 2294 LSPDVQSSTTPNNLSSESSTVE-----TPKTSSEVSLNSEEPSTTEAPTTLSLDILSTTN 2349  
QY 398 -----TKKPAPTAPK-----EPAPTTPKETAPTTPKKLTP 427  
Db 2350 NLSQSTVSTEDKSETSSENSESEKPT--SAPELVTSVTHVASSSPDPVPTES--SEPDLTG 2406  
QY 428 TTPKEKLAPTTPKEPAPTTPPEELAPTTPPEPTPTTPEP-----APTTPKAAAPN 476  
Db 2407 SSTENPEASSKQSTISSTPTPDTTASEEPTKSTSPDLSSTSNVLSSESTTPESS-- 2464  
QY 477 TPKEPAPT-----PKPEPAPT-----TPKEPAP-----TTPKETAPTTP 510  
Db 2465 --KSPVSSSTEGISVSTSTEFKSPVESTISSVLEEDLTKTTPSPILEETTTASETSEPLT 2522  
QY 511 KGTAPTTLKEPAPTTPPKPAPKELAPTTTKEPTSTTSDKPA-----PT----- 553  
Db 2523 EDLSLVSVRIHELTTSSENVPKESSTTTTSSSESKSPQEPAGILTSTVVVPTSSVSLITA 2582  
QY 554 -----TP--KGTAP--TTPK-----EPAPTTPKEPAPTTPKGTAPTTLKEPAPT 593  
Db 2583 SEIEAITSNTTPKQGRPTITTSKSLVKSTTSPSTVTSSEPSSESTKRTTVSTTVSTTPT 2642  
QY 594 TPKKPAPKELAPTTTKGPT--STTSDKPAPTTPKETAPTTPKEPAPTTPKPAITTPPETTP 652  
Db 2643 EETTTSESLILTAAPSKPTESTTESSEAPTPAKTSETKPSNVSSTSRKSTENVENTSQ 2702  
QY 653 PTTSEVSTPTTKEPTTIHKSPDESTP--ELSAEPTPKALE--NSPKPEGVPTTKTPAAT 708  
Db 2703 SCSLESTMSSTS-----SEPETNAPAVTVSSEASTTLEENSSSTSSP-----TSSEASV 2752  
QY 709 K-----PEMTTAKDKTTERDLRTTPTETTTAAPKMTKETATTTTEKTESKITATTT----- 759  
Db 2753 KLSSLPESITSEAVTVSSR-----APAEITMSSSEHIREISTVSSESEPEIPLSTTVSPN 2808  
QY 760 -----QVTSTTQDTP-FXITT-----LKTTLTLPKRVTTTKKITTITTEIMNKP 802  
Db 2809 VVTASSIPSEEPILSVSTSSSTPRVRLITGTDDLLIVSVTVPSHGNNRRONITASSV--P 2865  
QY 803 EETAKPKDRATNSKATTPKPKQRTKAPKAPTSTKKPKTTPRVRKPK-----TTPTPKMTST 859  
Db 2866 SNSTSPILPSES-LTTPQPPPTTTTAKPATTSKGRGPPPSIQPPAEEMFTTAP----- 2918  
QY 860 MPELN-----PTSRIAEAMLQTTT 878  
Db 2919 PPSNGGYGEETNQEEQVSTTTT 2942  
RESULT 15  
I51618  
nucleolar phosphoprotein - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 24-Sep-1999  
C:Accession: I51618; S57157  
R:Cairns, C.; McStay, B.  
J. Cell Sci. 108, 3339-3347, 1995  
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp  
A:Reference number: I51618; MUID:96019267  
A:Accession: I51618  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-990 <CAI>  
A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921  
C:Genetics:  
A:Gene: xNopp180  
A:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein  
C:Keywords: phosphoprotein  
Query Match 12.2%; Score 626.5; DB 2: Length 990;  
Best Local Similarity 26.2%; Pred. No. 4.9e-20;  
Matches 253; Conservative 132; Mismatches 393; Indels 187; Gaps 45;

```
QY 3 DNKKNRKTKKPKPPVVDGAGSLDNGDFKVTPDTSTTOHKNVSTSPKTIITAKPINR 62
Db 59 DAKKRPPANGLPKKSSSESSSEDEPPAKKRAQ-----PAGGKPPVWKAQPK 114
QY 63 PSLPPNSDTSKTSLVNKETTVE--TKETTNNK-----OTSDGKEKTSKAKETOSIEK 116
Db 115 KAKSSSEDSSESD--SEETKKPPAKRPAQTPKVAAVKTPQKKAKSSS-ESSSED 170
QY 117 TSAKDLAPTSKVLAKTPKAEITTKGPAITTKPEPTTPKE-----PASTTPKEPTPTI 172
Db 171 EASKKQPIKV-----PPQAVVWAGLAGNNGKTADSSSESDSPPAKKTAAKTPPT- 225
QY 173 KSAPTPKPEPAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTT--TKPEAPTT-----T 226
Db 226 --KPATAAKPOAKKTAGKSSSREDSDEQKTAQKPKPDVYSAVPPPTSVSKKKT 283
QY 227 KSAPTTKPEAPTT-----TPKKPAPTTKPEAPTT-----KEPTTPKPEAPTT 272
Db 284 LSQPGTKAKPESSDSDSDSEEQPAKAKIVPAKAAASAPKLAKKAETSTDSSESDSS 343
QY 273 KEAPTTKPEAPTAAPKKPAPTTKPEAPTT--KEPAP-----TT 311
Db 344 EDEKSSVKLGWKAAPK-APAAP--DAKSTPVAAAKKSAPAKKASSSDSDSSNEETT 400
QY 312 TKEPSPTTKPEAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEPSPTTKPEAPTTKPEPA- 370
Db 401 TKPAAKTTPAKSAATTSKTPTNGK--ATPTSKTPAKPGTKSTAKKSDSSSDSDSS 457
QY 371 ---PTTPKKPAPTTKPEAPTT--TPKEPAPTTTKKPAKPAKPEPA-PTTPKETA----- 418
Db 458 SDEETTKPAAKTTPAKSAATTSKTP--TNSKATPTSKKTPAKPGTKSAAKKSDSS 514
QY 419 -----PTTPKKLPPTPEKLPAPTTPEKAPTTPEELAPTTPEPTPTTPEEPAPTT-PA 472
Db 515 SDSSSDSSDEKTPA--KRAAKTTPAKA-----AKTTPAKA-----AKTTPAKP 558
QY 473 AAPNTKPEAPT---TPKEPAPTTKPEAPTTKPEAPTTKPEAPTT--KGTAPTTKPEAPTTPK 527
Db 559 AAKSTPGQVPTKKSSSDSDSDSEDEKKSAPKPAVKTTPGKATS-----KPWVAS 611
QY 528 KPAPKELAPTTTKPTSTSDKAPTTPKGTAPTTKPEAPTT--PKEPAPTTPKGTAPT 585
Db 612 KPVPK-----KASSSDSDSSEETTKTKPLTKLSPAVKTLPPKKAESSDSDSDSD 665
QY 586 TLKEPAPTTPKPAKPELAPTTTKGP-----TSTTSKAPAPTTPKETAPT--TPKEP 635
Db 666 SERK---TKPAKPPAKSATPVNTKAPQNKASKASCSDSDSDSSEEGKSKQPTGKSPA 722
QY 636 APTTPKK-PAPTTPEPTPTTSEVSTPTTKE-PTTIHKSPDESTPELSAEPKPALENS 693
Db 723 ATAPPKKNPVAVNKPSSSSSDSDSDDEKQKPKQAAAAKDVKGAKAAKPTPKKAASS 782
QY 694 KPEPGVPTTKPAATKPEMTTAKDPTTERDLRTTPTTETAAPKMTKETATTTTEKTESK 753
Db 783 SSE---DSSSEDEVSAKKTNTAVSKSPV---TTPAAVPAAKKESSESSSEDEKQGG 835
QY 754 ITATTQVSTSTTQDTPPEKTIITLTKTTLAPKVTTK--KTIITTIMNKPEETA-KPKD 810
Db 836 KNTSTTKIANST-----PKAAAECSESSSEDEGKANGTSGCKRKR 877
QY 811 RATNS---KATTPKOKPKAPKPTSTKPKTMRPRVRPKTPTTP-RKWTSTMPELNPT 866
Db 878 ESTGNAECEAVTPE-----NKKLAKASPNTPFPKVNKKELANTPFRVVEEDIEINP- 928
QY 867 SRIAE 871
Db 929 -RMAD 932
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:21 ; Search time 47.4 Seconds  
(without alignments)  
748.767 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1167  
Perfect score: 5155  
Sequence: 1 VKDNKKNRKTKKTPPPVV.....GKPDGLTLRLNGTLVAFRG 968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1168	22.7	5179	1 MUC2_HUMAN	Q02817; Q14878; PRT; 5179 AA.
2	942	18.3	1664	1 SLP1_CLOTH	Q06852 clostridium
3	763	14.8	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	12.6	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	556.5	10.8	875	1 FPL_MYTE	Q25460 mytilus edu
6	555.5	10.8	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
7	551	10.7	620	1 NFB_MOUSE	P13983 nicotiana t
8	533	10.3	1087	1 NFB_MOUSE	P19246 mus muscucu
9	530.5	10.3	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	10.3	865	1 CPN_DROME	Q02910 drosophila
11	518.5	10.1	872	1 FPL_MYTE	Q25434 mytilus cor
12	509.5	9.9	662	1 MUC1_XENLA	Q05049 xenopus lae
13	503.5	9.8	1970	1 RPB1_HUMAN	P24928 homo sapien
14	498.5	9.7	1970	1 RPB1_MOUSE	P08775 mus muscucu
15	497	9.6	831	1 NPH_RAT	P16884 rattus norv
16	493.5	9.6	467	1 RPB1_CRIGR	P11414 cricetus
17	488.5	9.5	826	1 SSP2_PLAYO	Q01443 plasmodium
18	488	9.5	1020	1 NFB_HUMAN	P12036 homo sapien
19	475.5	9.2	267	1 EXTN_MAIZE	P14918 zea mays (m
20	471.5	9.1	5376	1 ZAN_MOUSE	O88799 mus muscucu
21	468.5	9.1	634	1 HWPL_CANAL	P46593 candida alb
22	467	9.1	817	1 VRP1_YEAST	P37370 saccharomyc
23	454	8.8	797	1 VGLX_HSVB	P28968 equine herp
24	454	8.8	1161	1 Y9P_YEAST	P47179 saccharomyc
25	453.5	8.8	2142	1 BAT2_HUMAN	P48634 homo sapien
26	448.5	8.7	670	1 V650_HSV1	Q00130 ictaluriid h
27	442.5	8.6	751	1 FPL_MYTE	Q27409 mytilus gal
28	439.5	8.5	1083	1 T2D3_HUMAN	O00268 homo sapien
29	432.5	8.4	439	1 XP2_XENLA	P17437 xenopus lae
30	432	8.4	3164	1 TEGU_HSV1	P10220 herpes simp
31	426	8.3	2715	1 TRX2_HUMAN	Q9um66 homo sapien
32	424.5	8.2	1125	1 MAP4_MOUSE	P27546 mus muscucu
33	424	8.2	307	1 SGS3_DROME	P02840 drosophila

## ALIGNMENTS

RESULT	1				
MUC2_HUMAN					
ID	MUC2_HUMAN	STANDARD;			
AC	Q02817; Q14878;				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).				
GN	MUC2 OR SMUC.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine;				
RX	MEDLINE=94132002; PubMed=8300571;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;				
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.				
RT	Identification of the amino terminus and overall sequence similarity				
RT	to prepro-von Willebrand factor."				
RL	J. Biol. Chem. 269:2440-2446(1994).				
RN	[2]				
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.				
RC	TISSUE=Colon;				
RX	MEDLINE=93016075; PubMed=1400449;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,				
RA	Kim Y.S.;				
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located				
RT	both upstream and downstream of its central repetitive region."				
RL	J. Biol. Chem. 267:21375-21383(1992).				
RN	[3]				
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.				
RX	MEDLINE=9135817; PubMed=1885763;				
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,				
RA	Petersen G.M., Kim Y.S.;				
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays				
RT	and polymorphism."				
RL	J. Clin. Invest. 88:1005-1013(1991).				
CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND				
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A				
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS				
CC	AGENTS AT MUCOSAL SURFACES.				
CC	- SUBUNIT: MULTIMERIC.				
CC	- SURCELLULAR LOCATION: SECRETED.				
CC	- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,				
CC	BRONCHUS, CERVIX AND GALL BLADDER.				
CC	- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR				
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).				
CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND				
CC	VARIES AMONG DIFFERENT ALLELES.				
CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT				
CC	OF SILKWORM HEMOCYTIN.				
CC	- SIMILARITY: CONTAINS 2 VWFC DOMAINS.				
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).				

Q09550 caenorhabdi  
Q10172 schizosacch  
P28955 equine herp  
Q28983 sus scrofa  
P54259 homo sapien  
P34926 rattus norv  
Q9V2n3 homo sapien  
P03200 epstein-bar  
P54258 rattus norv  
P46013 homo sapien  
Q9Y618 h nuclear r  
Q113428 homo sapien

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; L21928; AAB95295.1; -  
DR EMBL; M74027; AAA59875.1; -  
DR EMBL; M94131; AAA59163.1; -  
DR EMBL; M94132; AAA59164.1; -  
DR MIM; 158370; -  
DR InterPro; IPR000359; Cys\_knot.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002400; GF\_cysknott.  
DR InterPro; IPR001007; WFEC.  
DR InterPro; IPR001846; Vwd.  
DR Pfam; PF00007; Cys\_knot; 1.  
DR Pfam; PF00094; Vwd; 4.  
DR PRINTS; PR00438; GFCYSKNOT.  
DR SMART; SM00214; VWC; 2.  
DR SMART; SM00011; VWC\_def; 2.  
DR SMART; SM00216; VWD; 4.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE; PS01185; CTCK\_1; 1.  
DR PROSITE; PS01225; CTCK\_2; 1.  
DR PROSITE; PS01208; WFC; 2.  
KW Glycoprotein; Repeat; Signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 5179 MUCIN 2.  
FT DOMAIN 1401 1747 APPROXIMATE REPEATS.  
FT REPEAT 1401 1416 1.  
FT REPEAT 1417 1432 2.  
FT REPEAT 1433 1448 3.  
FT REPEAT 1449 1464 4.  
FT REPEAT 1465 1471 5.  
FT REPEAT 1472 1478 6.  
FT REPEAT 1479 1494 7A.  
FT REPEAT 1495 1517 7B.  
FT REPEAT 1518 1533 8A.  
FT REPEAT 1534 1556 8B.  
FT REPEAT 1557 1572 9A.  
FT REPEAT 1573 1596 9B.  
FT REPEAT 1597 1612 10A.  
FT REPEAT 1613 1635 10B.  
FT REPEAT 1636 1651 11A.  
FT REPEAT 1652 1675 11B.  
FT REPEAT 1676 1693 12.  
FT REPEAT 1694 1699 13.  
FT REPEAT 1700 1715 14.  
FT REPEAT 1716 1731 15.  
FT REPEAT 1732 1747 16.  
FT DOMAIN 4815 4886 WFC 1.  
FT DOMAIN 4924 4991 WFC 2.  
FT DOMAIN 5075 5160 CTCK.  
FT DISULFID 5075 5122 BY SIMILARITY.  
FT DISULFID 5089 5136 BY SIMILARITY.  
FT DISULFID 5098 5152 BY SIMILARITY.  
FT DISULFID 5102 5154 BY SIMILARITY.  
FT DISULFID ? 5159 BY SIMILARITY.  
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 770 770 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 894 894 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 1351 1351 H -> L (IN REF. 3).  
FT CONFLICT 1412 1412 T -> S (IN REF. 3).  
FT CONFLICT 1449 1449 L -> P (IN REF. 3).  
FT CONFLICT 1504 1504 M -> T (IN REF. 3).  
FT CONFLICT 4192 4192 G -> S (IN REF. 2).  
SQ SEQUENCE 5179 AA: 540295 MW: 85CD7571FB9A5663 CRC64;  
  
Query Match 22.7%; Score 1168; DB 1; Length 5179;  
Best Local Similarity 32.5%; Pred. No. 4.8e-39;  
Matches 363; Conservative 53; Mismatches 448; Indels 254; Gaps 40;  
  
QY 21 DEAGSLDNGDFK-----VTTPTDSTTQH-NKVTSPK----- 52  
DB 1312 DHPSSGSDGDRPFDGVCQGPEDIECRSVKDPHLSLEHQGVQDVSVGFICKNEDQF 1371  
QY 53 -----ITTAKPINPRPSLPPNSDTSKETSIVNKTETVEIK 88  
DB 1372 GNGPFGLCYDYKIRVNCWCPMDKCIITPSPPTTPSPPTTTTLLPTTTPSPPTT----- 1427  
QY 89 ETTTNKQISTDGEKKTSAKETQSIKTSKADLAPTSKVLAKPTKAEITTKGPAITTP 148  
DB 1428 -TTTTPPTTTPSPPTTTTTP-----LPTT-----TPSPPISTTTTTPPTTTP 1470  
QY 149 KEPT-----PTTPKEPASTTPKEPTTTIKSAP-TTPKEP-----APTAKSAP-TTP 194  
DB 1471 SPPTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPSPPTTTP 1530  
QY 195 KEPAITTKKEPAPTTPKEP-----APTTPKEPAPTTPKSAAP--TTPKEPAPTTPKKA 245  
DB 1531 TTPPTTTPSPPTTTPITPTTSTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPSPPT 1590  
QY 246 PTPKEPAPTTPKEPTTTPKEPAPTTPKEPAPTTPKEP-----APTAKKPAITTPKE 298  
DB 1591 ITTTTPPTTTPSPPTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1650  
QY 299 PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTTPKEPAPTTPKSAAP---TTPKEPSP 355  
DB 1651 PPPTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPTTTP 1709  
QY 356 TTTTPKEPAPTTPKEPAPTTPKPAITTPKEPAPTTPKEPAPTTPKPAITTPKEPAPTTP- 414  
DB 1710 TTTTTPSSSTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1769  
QY 415 ----KETAPTTP-----KKTLP 427  
DB 1770 FSPFSTTTTTPCVPLCNWTGMLDSGKPNFKPGGDTLIGDVCQGPWAANTSCRATWTP 1829  
QY 428 -----DVPIGOLGOTVVCDVSVGLICKNEQKPGVIPMAFCLNVEINVOCCVOTPTTMTT 1889  
DB 1830 DVPIGOLGOTVVCDVSVGLICKNEQKPGVIPMAFCLNVEINVOCCVOTPTTMTT 1889  
QY 438 PKKPAPTTPPELAPTTPPEPTPT-TPEEPAPTTPKAAANTPKAPAPTTPKEP-----AP 491  
DB 1890 TENPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1949



```
QY 871 EAMLOTTTRPNOTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPMDYLP 927
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D-CRC64;

Query Match 14.8%; Score 763; DB 1; Length 1367;
Best Local Similarity 28.5%; Pred. No. 1.le-23;
Matches 286; Conservative 98; Mismatches 463; Indels 154; Gaps 43;

QY 35 TTPDTSTTOHNKVSSTPKITTAAPINP-----RPSLPPNSDTSKETSITVNKETTVE 86
Db 236 STSESSTTSSTSESSTSSSTTAPATPTTCTKEKPTPTTCTKEKPTPHHDTPC 295
QY 87 TKETTTNKQSTDKKETSQAKETQSEKTSKADLAPTSKVLAKPTPKAETT--TKGPA 144
Db 296 TKKTTTSK-TCT--KKTTPVTPS-SSITSSAPV-----PTSSSTSSAPV 344
QY 145 LTPKE-----PTTPKPEASTTPKEPTTIKAP-----TTPKEPAPTTPKSAPTTP 194
Db 345 TSSTSSAPVPTPSSSTSSAPVTSSTSSAPVTSSTSSAPVPTPSSSITE 404
QY 195 KEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAP 254
Db 405 SSSAPVTS-----TTESSAPVTS--STTESSAPVT--SSITSSAPVTSSTES 453
QY 255 TTPKEPTP---TTPKEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAP 307
Db 454 SSAPVPTPSSSTSSAPVT---SSTSSAP-VPTPSSSTSSAPVPTPSSSTESS 509
QY 308 APTTKPEPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAP 363
Db 510 APVPTPS-SSTTESSAPAPTSSSTSSAPVTSSTSSAPVTSSTSSAPVPTPSSSTESS 568
QY 364 ---TTPKEPAPT-TPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKETA 418
Db 569 VTSSITSSAPVPTPSSSTSSAPVPT---PSSITSSAPAPTSSITSSAPVPTSSITSSA 624
QY 419 PTPPKLPTTPKEKLAAPTTPKPEAPTTPPELAPT-TPEEPTTPPEAP-----TTPKA 472
Db 625 PVT---SSTTESSAP-VPTPSSSTSSAPVPTPSSSTSSAPVPTPSSSTSS 679
QY 473 AAPNTKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT-TPKGAPTTPKPEAPTTPKPKAP 531
Db 680 SAPVT---SSTTESSAPVT---SSTTESSAPVPTPSSSTSSAPVPTPSSSITE 731
QY 532 KELAPTTPKEPTSTSDKPAPTTPKGTAPTTPKPEAP---TTPKEAPT-TPKGAPTTL 587
Db 732 SSSAPVPTPS-SSTTESSAPVTSSTSSAPVPTPSSSTSSAPVPTPSSSTESS 790
QY 588 KEPAPTTPKPKAPKELAPTTPKTGPTSTSDKPAPTTPKGTAPTTPKPEAP-----TTPKK 642
Db 791 SAPVPTPSSSTSSAPVPTPSSSNITSSAPSSITSSAPSSITSSAPVPTPSSSTESS 850
QY 643 PAPTPPETPPTTSEVSTPT-----TKETTTIHKSPDESTPELSAETPKALENSKPEP 697
Db 851 SAPVSSSTTESSAPVPTPSSSNITSSAPSSI---PFSSTTESFTGT-TVTPSSSKVP 906
QY 698 GVPTTKPAATKPEMTTAKDKTERDLRTPTTTTAAAPKMKETATTTTEKITESITAT 757
Db 907 GSQTETSVSSTTETTVPTTKTTTSVTTPTTTTTCSTGTSNAGETSGGSPKVTVT 966
QY 758 --TTOVTSTTTODTTPFKLITLTKTTTLAPKVT---TKKTTT-BIMNKEPEATAKPKDR 811
Db 967 VPTTTTTSVTSTTTTITTCSTGTSNAGETSGGSPKTIITTVPCSTSPSEFA-----1021
QY 812 ATNSKATTPKPKOP-----TKAPKPTSTKPKMTPRVKPKPTTPTPR 854
Db 1022 ---SESTTSPPTPVTTVVSTTVVTTTEYSTKPGGEITTTFTVKNIPTTLTIAPT-1077
QY 855 KMTSTPELNPSTRIAEAMLQTTTRNQTPNSKLVNPKSDEAGAGETPHMLLRPHV 914
Db 1078 -SVTTVTFNFTPTT-ITTVTCSTGT-----NSAGETSGCSPKVTITVPCST 1122
QY 915 FMEVTPMDYLPVNPNOGIIINPMLSDETNICNGKPVVDGLTT 957
```

Db 1123 GRGEVTEATTLVTAVTITVTTESSTGNSA-GKTTTGYTT 1164

RESULT 4

MUC1\_HUMAN

AC MUC1\_HUMAN STANDARD; PRT: 1255 AA.

ID P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)

DE (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)

DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-

DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN

DE DF3).

GN MUC1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Pancreas;

RX MEDLINE=90368716; PubMed=2394722;

RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;

RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.;"

RL J. Biol. Chem. 265:15294-15299(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=90202794; PubMed=2318825;

RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;

RT "Epistatin, a carcinoma-associated mucin, is generated by a

RT polymorphic gene encoding splice variants with alternative amino

RT terminl.;"

RL J. Biol. Chem. 265:5573-5578(1990).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE-Breast carcinoma;

RX MEDLINE=90368715; PubMed=1697589;

RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,

RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;

RT "Molecular cloning and expression of the human tumor-associated

RT polymorphic epithelial mucin.;"

RL J. Biol. Chem. 265:15286-15293(1990).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=91097524; PubMed=2268309;

RA Lancaster C.A., Peat N., Duhig T., Wilson D.,

RA Taylor-Papadimitriou J., Gendler S.J.;

RT "Structure and expression of the human polymorphic epithelial mucin

RT gene: an expressed VNTR unit.;"

RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).

RN [5]

RP SEQUENCE FROM N.A.

RC TISSUE-Breast carcinoma;

RX MEDLINE=90276413; PubMed=2351132;

RA Wreschener D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,

RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,

RA Keydar I.;

RT "Human epithelial tumor antigen cDNA sequences. Differential splicing

RT may generate multiple protein forms.;"

RL Eur. J. Biochem. 189:463-473(1990).

RN [6]

RP SEQUENCE FROM N.A.

RC TISSUE-Breast carcinoma;

RX MEDLINE=90276414; PubMed=2112460;

RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,

RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschener D.H.;

RT "A transcribed gene, containing a variable number of tandem repeats,

RT codes for a human epithelial tumor antigen. cDNA cloning, expression

RT of the transfected gene and over-expression in breast cancer

RT tissue.;"

RL Eur. J. Biochem. 189:475-486(1990).

RN [7]

RP SEQUENCE FROM N.A.

RX MEDLINE=91033045; PubMed=1688329;

RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,

RA Weitsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschener D.H.;

RT "Isolation and characterization of an expressed hypervariable gene

RT coding for a breast-cancer-associated antigen.;"

RL Gene 93:313-318(1990).

RN [8]

RP PARTIAL SEQUENCE FROM N.A.

RX MEDLINE=88330762; PubMed=3417635;

RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,

RA Burchell J.;

RT "A highly immunogenic region of a human polymorphic epithelial mucin

RT expressed by carcinomas is made up of tandem repeats.;"

RL J. Biol. Chem. 263:12820-12823(1988).

RN [9]

RP SEQUENCE OF 1-169 FROM N.A.

RX MEDLINE=90088473; PubMed=2597151;

RA Abe M., Siddiqui J., Kufe D.;

RT "Sequence analysis of the 5' region of the human DF3 breast

RT carcinoma-associated antigen gene.;"

RL Biochem. Biophys. Res. Commun. 165:644-649(1989).

RN [10]

RP SEQUENCE OF 1-109 FROM N.A.

RC TISSUE-Thyroid;

RX MEDLINE=96183746; PubMed=8608966;

RA Weiss M., Baruch A., Keydar I., Wreschener D.H.;

RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse

RT transcriptase polymerase chain reaction of the MUC1 gene.;"

RL Int. J. Cancer 66:55-59(1996).

RN [11]

RP SEQUENCE OF 1-89 FROM N.A.

RC TISSUE-Lung;

RX MEDLINE=96181716; PubMed=8604237;

RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,

RA Lee L.N., Luh K.T., Wu C.W.;

RT "Mucin mRNA expression in lung adenocarcinoma cell lines and

RT tissues.;"

RL Oncology 53:118-126(1996).

RN [12]

RP SEQUENCE OF 1-46 FROM N.A.

RC TISSUE-Breast carcinoma;

RX Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;

RA Submitted (OCT-1992) to the EMBL/Genbank/DBJ databases.

RL -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN

CC CYTOSKELETON.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM

CC IS ALSO PRODUCED.

CC -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE

CC SPLICING.

CC -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL

CC TUMORS, SUCH AS BREAST CANCER.

CC -!- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND SIALIC

CC ACID).

CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT

CC VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE

CC MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.

CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; J05582; AAA60019.1; -

CC EMBL; M32738; AAA35804.1; -

CC EMBL; M32739; AAA35806.1; -

CC EMBL; J05581; AAA59876.1; -







CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.

CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE

CC ZONA PELLUCIDA.

CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,

CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF

CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHUS.

CC -1- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT

CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).

CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.

CC -1- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF0533356; AAC78790.1; -.

CC EMBL; U83191; AAC51208.1; -.

CC MIM; 602372; -.

CC InterPro; IPR000561; EGF-like.

CC InterPro; IPR000998; MAM.

CC InterPro; IPR002965; P\_rich\_extensn.

CC InterPro; IPR002919; Tila.

CC InterPro; IPR003328; Tila.

CC InterPro; IPR001007; VWFC.

CC InterPro; IPR001846; VWFD.

CC Pfam; PF00629; MAM; 3.

CC Pfam; PF01826; TIL; 4.

CC Pfam; PF02345; Tila; 4.

CC Pfam; PF00094; vwd; 4.

CC PRINTS; PR01217; PRICHEXTENS.

CC SMART; SM00137; MAM; 2.

CC SMART; SM00214; VWC; 1.

CC SMART; SM00011; VWC def; 3.

CC SMART; SM00216; VWD; 3.

CC PROSITE; PS01186; EGF\_2; 3.

CC PROSITE; PS00740; MAM\_1; 1.

CC PROSITE; PS00600; MAM\_2; 4.

CC Glycoprotein; Transmembrane; Cell adhesion; Repeat.

CC NON\_TER 1

CC DOMAIN 1

CC DOMAIN 2 (PARTIAL).

CC DOMAIN 3

CC DOMAIN 4

CC 66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)

CC (MUCIN-LIKE DOMAIN).

CC VWFD 1 (PARTIAL).

CC VWFD 2.

CC VWFD 3.

CC VWFD 4.

CC VWFD 5.

CC EGF-LIKE.

CC ?

CC CARBOHYD 74

CC CARBOHYD 403

CC CARBOHYD 1023

CC CARBOHYD 1099

CC CARBOHYD 1618

CC CARBOHYD 1737

CC CARBOHYD 1832

CC CARBOHYD 1878

CC CARBOHYD 2136

CC CARBOHYD 2505

CC CARBOHYD 2374

CC CARBOHYD 2700

CC NON\_TER 2700

CC SEQUENCE 2700 AA; 293013 MW; 80E60C0B12277B1 CRC64;

Query Match 10.8%; Score 555.5; DB 1; Length 2700;

Best Local Similarity 32.6%; Pred. No. 2.3e-15;

Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;	
QY 58	PINRPSLPNSDTSKETSIVNKKETIVETKETTITNKQTSDGKEKTSKETSIEKT 117
Db 445	PVKVLPPLPPVSPVSS-----TGPSETTGLTENPSTK-----KPTVSIK 487
QY 118	SAKLAPTSKVLAKPTPKAETTTGPAULTTTPKEPTTPPKEPASTTPKEPT-----PTTI 172
Db 488	SVTTEKPT-----VPKEKPTTPTKPTISTEK---PTIPSEKPNMPSEKPIPESEKPTIL 539
QY 173	KSAPTTKEPAPTTTKSAPTTKPEAPTTTTPKEAPTTTTPKEAPTTTTPKEAPTTTTPK 231
Db 540	TEKPTIPSE-KETIPSEKPTISTEKPTVPTTEB--PTTTEETTYMEEPVPTKPSIPT 596
QY 232	TPKEAPTTTPKAPTTTPKEAPTTTPKEPTTPTPKEAPTTTPKEAPTTTPKEAPTPAKKP 291
Db 597	--EKSPITEK---PTISMEETIISTEKPTICPEKPTIPTK---PTIPTKSTISPEK- 647
QY 292	APTTTPKEAPTTTPKPEAPTTTPKPSPTTPKEAPTTTTSAPTTTTPKEAPTTTTPKAPT 351
Db 648	-PTTPTPE-KPTIPTKPTISTEKPTIPTK-PTISPEKLTIPTEKLTIPT---EKPTIPT 701
QY 352	EPSPTTTPKEAPTTTPKPEAPTTTPKPA-----PTTPKPEAPTTTPKEAPTTTTPKAPTAP 406
Db 702	EKPTISTEE--PTTPTTEETIISTEKPSIPMEKPTIPTTEETTSVEETIISTEKLTIPM-- 757
QY 407	KEPAPTTPKETAPT-----TPKKLTTPTEKLA-----PTTPKEAPTTPELAPTTPPE 456
Db 758	EKPTISTEKPTIPTKPTISTEKLTIPT-EKLTIPTEKPTIPIEETIISTEKLT--TIPT 814
QY 457	PTPTTPKEAPTTTPKAAAPNTKPEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKGTAPT 516
Db 815	KETISPEKPTISTEK-----PTIPTK-PTIPTPE-----ETTISTEKLTIPT 855
QY 517	TLKEPAPTTPKKPAKELAPTTTKEPTSTT-----SDKPAPTTPKGTAPTTPKEAPTT 570
Db 856	--EKPTISPEKLTIPTEKPTISTEKPTIPIEETIISTEKLTIPTEKPTIPT--EKLALR 911
QY 571	PKPEAPTTTPKGTAPTTLKPEAPTTTPKKAPKELAPTTTTPKGTSTTSDDKAPTTPKET--A 628
Db 912	PPHPSPETA-TGLAALVMSHPASPMTSV---ILGTTTTSRSTGMSMCP-PNARYESCAC 966
QY 629	PTTPKEAPTT 638
Db 967	PASCKSPRPS 976
RESULT 7	
EXTN_TOBAC	
ID	EXTN_TOBAC STANDARD; PRT; 620 AA.
AC	P13983;
DT	01-JAN-1990 (Rel. 13, Created)
DT	01-JAN-1990 (Rel. 13, Last sequence update)
DT	01-MAR-1992 (Rel. 21, Last annotation update)
DE	EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
GN	HRGPNT3.
OS	Nicotiana tabacum (Common tobacco).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC	Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX	NCBI_TaxID=4097;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. XANTHI; TISSUE=Leaf;
RX	MEDLINE=90128263; PubMed=2612909;
RA	Keller B., Lamb C.J.;
RT	"Specific expression of a novel cell wall hydroxyproline-rich
RT	glycoprotein gene in lateral root initiation.";
RL	Genes Dev. 3:1639-1646(1989).
CC	-1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC	THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC	MAIN ROOT.
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.



DR Pfam: PF00038; filament; 1.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
 KW Repeat.  
 FT DOMAIN 1 97 HEAD.  
 FT DOMAIN 98 408 ROD.  
 FT DOMAIN 409 1087 TAIL.  
 FT DOMAIN 436 517 GLU-RICH (ACIDIC).  
 FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.  
 FT DOMAIN 887 1087 GLU/LYS-RICH.  
 FT DOMAIN 98 129 COIL 1A.  
 FT DOMAIN 130 141 LINKER 1.  
 FT DOMAIN 142 239 COIL 1B.  
 FT DOMAIN 240 261 LINKER 12.  
 FT DOMAIN 262 283 COIL 2A.  
 FT DOMAIN 284 287 LINKER 2.  
 FT DOMAIN 288 408 COIL 2B.  
 FT CONFLICT 133 133 K -> QA (IN REF. 2 AND 3).  
 FT CONFLICT 199 199 A -> AR (IN REF. 2 AND 3).  
 FT CONFLICT 281 281 S -> T (IN REF. 2 AND 3).  
 FT CONFLICT 492 492 L -> G (IN REF. 2 AND 3).  
 FT CONFLICT 551 551 P -> PREAKSP (IN REF. 3).  
 FT CONFLICT 689 712 MISSING (IN REF. 3).  
 FT CONFLICT 714 714 G -> A (IN REF. 3).  
 FT CONFLICT 814 814 V -> M (IN REF. 2 AND 3).  
 FT CONFLICT 843 843 T -> N (IN REF. 2 AND 3).  
 SQ SEQUENCE 1087 AA; 116612 MW; 57BAC76A38EDICB9 CRC64;  
 Query Match 10.3%; Score 533; DB 1; Length 1087;  
 Best Local Similarity 27.5%; Pred. No. 8.7e-15;  
 Matches 216; Conservative 81; Mismatches 330; Indels 158; Gaps 38;  
 QY 21 DEAGSLDNGDFKVT-----TDTSTQHNKYSTSPKTTAKPINRPSLPSNDSITSKET 75  
 DB 404 EECRIGFGSPSLTBGLPKIPSIST--HIKVKSEMIKWVE-----KSEKET 449  
 QY 76 SLTVNKETIVETKTTNKQISTDQKETTSAKETQSTIEKTSKADLAPTSKVLAKPTPK 135  
 DB 450 VIVEGQTEIRVTEGTEEDKEAQOGEAEAEGEREEELAAATSPPAEEAASPEKE 509  
 QY 136 AETTTKGPALTTPKEPTTPPTTPKEPASTTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTPK 195  
 DB 510 TKSRRVEAKSPGEAKSPGEAKSPA-----EAKSPGEAKSPGEAKSPAEPKSPA 564  
 QY 196 EP-APTTPKEPAPTTTPKEPAPTTTPKEP-----APTTPKSAPTTPKEPA-----PTTPKKPA- 245  
 DB 565 EPKSPAEAKSPA--EPKSPA--TVKSPGEAKSPSEAKS--PAKSPAEAKSPAESPAE 619  
 QY 246 ---PTTPKEPA-----PTTPKEP-----TPTTPKEPAPTTTPKEPAPTTTPKEPA-----PTAPKK 290  
 DB 620 AKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 676  
 QY 291 P-----APTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 342  
 DB 677 PGEAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 729  
 QY 343 TKSAPTTPKE--PSPTTPKEPA-----PTTPKEPAPTTTPKPKPAPTTTPKEPAPTTTPKEPAPTT 397  
 DB 730 SPAVKSFGKSPGKSPGKSPGKSPGKSPGKSPGKSPGKSPGKSPGKSPGKSPGKSPGKSPGK 789  
 QY 398 TKKPAFTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 457  
 DB 790 AKSPVKEDIPPAEAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 848  
 QY 458 TPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 505  
 DB 849 RP--PEQVKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 904  
 QY 506 APTTPKGTAPTTTPKEPAPTTTPKPKPAPKELAPTTTPKSTSTSD--KPAPTTTPKGTAPTTTP 563  
 DB 905 TLTPKTEAKSKDEAPKAPKPKVEKKEKTTPTTEKPKDSTAEAKKEAGEKKKAVASEE 964  
 QY 564 KEPATTPKPEAPPTPKGTAPTTLKEPAPTTTPKPKPAPKELAPTTTPKGTTPTSKDPAPT 623  
 DB 965 ETAPKLGKVEEA--KPKETETTKTEAEDTKAKEPS-----KPTETEKP----- 1006  
 QY 624 PKETAPTTKPEAPPTPKPKPAPTTTPPTTSEVSTPTTTPKPTTHKSPDSESTPELSA 683  
 DB 1007 -----KKEMPAAPEK-----KDKKEKTTESRKPPEK----- 1034  
 QY 684 EPTPKALENSPKPGVPTTKTTPAATKPEMTTAKDTRTTERDLRTTPTTTTAAAPKMTKETA 743  
 DB 1035 ----PK-MAEKVEDDKSLSEK--SKPKTEKAEKSSSTDQKESQPP-----KTTEDKA 1082  
 QY 744 TTTEK 748  
 DB 1083 TKGEK 1087  
 RESULT 9  
 TCNA\_TRYCR STANDARD; PRT; 1162 AA.  
 ID TCNA\_TRYCR  
 AC P23253;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE STALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).  
 GN TCNA.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID:5693;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-SILVIO X-10/4;  
 RC MEDLINE=91277609; PubMed=1711561;  
 RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,  
 RA Prioli R.P.;  
 RT "The Trypanosoma cruzi neuraminidase contains sequences similar to  
 RT bacterial neuraminidases, WTD repeats of the low density lipoprotein  
 RT receptor, and type III modules of fibronectin.";  
 RL J. Exp. Med. 174:179-191(1991).  
 RN [2]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=91376547; PubMed=1896773;  
 RA Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;  
 RT "Trypanosoma cruzi: localization of neuraminidase on the surface of  
 RT trypomastigotes.";  
 RL Trop. Med. Parasitol. 42:146-150(1991).  
 CC -!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN  
 CC PARASITE INVASION OF CELLS.  
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC  
 CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLNEURAMINYL  
 CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED  
 CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,  
 CC GLYCOLIPIDS OR COLOMINIC ACID.  
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
 CC (POSSIBLE).  
 CC -!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM  
 CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.  
 CC -!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT  
 CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TCNA  
 CC PROTEIN.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.  
 CC -!- SIMILARITY: CONTAINS 3 BNR REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: M61732; AAA30255.1; -  
 DR FIR: JH0557; JH0557.



J. Mol. Evol. 43:348-356(1996).

1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.

1- SUBCELLULAR LOCATION: SECRETED.

1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.

1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.

1- PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D63777; BAA09850.1; InterPro: IPR002964; Adhesive\_plaq. InterPro: IPR002965; P\_rich\_extensn. PRINTS; PR01216; ADHESIVEI. PRINTS; PR01217; PRICHEXTENS. SIGNAL; Repeat; Hydroxylation. CHAIN 1 20 POTENTIAL. FT CHAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN. FT DOMAIN 21 41 NONREPEAT LINKER. FT DOMAIN 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-P-[PST]-[ST]. FT DOMAIN 184 192 NONAPEPTIDE 1. FT DOMAIN 213 221 NONAPEPTIDE 2. FT SEQUENCE 872 AA; 101677 MW; 98CC70D/C75FF3C4 CRC64;

Query Match 10.1%; Score 518.5; DB 1; Length 872; Best Local Similarity 29.2%; Pred. No. 2.7e-14; Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

QY 19 VYDEAGSLDNGDFKVTTPDST--TOH-----NKVSTSPKITTAKPINRPSLPNS 69  
Db 29 VYGSAYSGASAGAYK-TLPGSHPGYGVKHPVYKPMNKIPT-PYI--SKKSYPAKYKGY 84  
QY 70 DTSKETSILVNKETTIVETKETTITNKOTSDGKETTSAKETQSIETKSADLAPTSKVL 129  
Db 85 YPTKRYQPTGYGKTNVPIYKPIAKLLSYKAIKITYPAYKAKTSYPPSYK-----HKIT 139  
QY 130 AKPTPKAETTTKGPALTTPEKEPT-PTTPKEPASTTPKEPTPTTIKSAPTTPKEPATTTK 188  
Db 140 YPTTYK-----PKITYP--PTYKQPSYPPSYKPKTTPPTTYK-----PKITYPPTYK 185  
QY 189 SAPT-TPKEPAPT--TTKEPAPTTPKEPAPTTPKEPATTTKSAPTTPKEPATT--PKK 243  
Db 186 RKPSYTPYKPKATYPPTYKPKITYP---PTYKPKPSYT-----PTYKPKPSYT 236  
QY 244 PAPTTPKEPAPT--PKETPTTT--PKEPAPTTPKEPAPTTPKEPAPT-APKKPAPTTP 296  
Db 237 SYPSIYKPKASYVSSYKSKKTYPTVKIYSYPTTKPKPSYP---PTYKPKVTPPTY 292  
QY 297 KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPT-----TTKEPATTTKSAPTTK 351  
Db 293 K-PKPSYPTPKKITYPPT-YKPKPSYTPYKQKPSYPPYKSKSYPTSYKSKKTYPP 350  
QY 352 EPSPTTKPEA-PTTPKEPAPTTPKPP-APT-TPK-----EPAPTTPKEPATTTKPP 401  
Db 351 TYKPKITYPTTPYKPKPSYPPSYKPKKITYSTYKPKITYPTTPYKPKSYPPSYKPKITY 410  
QY 402 APTAPKEPAPTTPKETAP-----TTPKLTPTTPPEKAPATTTPKEPAPTTPPEELAPTTPEE 456  
Db 411 T-YKPKISYPTTPYKPKASYVSSYKSKKTYPTTPYKPKISYPTTPYKPKPSYPPSYKPKITY 469  
QY 457 PT-PTTPPEEAPPTPKAAAPTTPK-EPAPTTPKEPAPTTP-----EPAPTTP--PKETAP 507

117 PVAATPTVQVIAAPVIAATPPVAAASAPTAAVTPVISPVPVPPVVPANTTPVFAAPV 176  
QY 280 PKPA--PTAPKPAATTPKPAATTPKPAATTPKKE--PSPTTPKEAP-----TTKSA 331  
Db 177 AAVPAAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 236  
QY 332 PTTTKEP---APTPTTKSAPTTPKESPTTTPKEA-PTTPKEPAPTTPKPKAPPTTPKEAP 387  
Db 237 PLAAAEPPVPAATETPPVPAAPAAASHVSAVPAVETAVAPVSAETPPVAAATLTAP 296  
QY 388 TTP-----KEPAPTTPKPAATTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 428  
Db 297 ETAPALPVVAESQVAANTVWAPPTTPPAPEPETIAPVVAETPEVASVAETTPVVPV 356  
QY 429 TPEKL-APTTPKEPAPTTPPEELAPTTPPEPTTPTEEP---APTTPKAAANTTPKEAPT 484  
Db 357 AAESIPAPVWATTPVAT---LAVTDPDVTSASAPVLPVPIAPSPVSAVETPVDLAP 413  
QY 485 TPKEPAPTTPKPAATTPKPAATTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPK 533  
Db 414 V-----LPPVAAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 469  
QY 534 LAPTTPKEPTSTSDKPAATTPKGTAPTTP-----KEPAPTTPKAPKAPKAPKAPKAP 585  
Db 470 AAPVSTTP--TASVPETTPAAVPTPEPIDVSVLSEAAIETPVAPVPEVTEVAVADV 527  
QY 586 TLKEPA-----PTTPKPKAPKELAPTTP-----KGTPTSTSDKPAATTP-----PKETAPT 531  
Db 528 APPEAAADLIIEVPEPPAPIPDLLEQTSVPAVEAAESTSS--PIPETSLPPNEAVASP 585  
QY 632 PKPAPTTPKPAATTPPTTPPTTSEVSTPTTKETPTTIHKSPD-----ESTPBL 681  
Db 586 EVAVAPITAEPTP-----BPEPLAPTEPIPEVPAVWVQEADEAVDEVPVETSTSIPT 641  
QY 682 SAEPTPKALENSKPEGVPTTPKPAATKPEMTTAKDKTTTERDLRTP--ETTAAPKMT 739  
Db 642 TVE-FPEAVAEKVLDPAL--TEAPVTTPQEDVANINDGAPATEI-TTPAVEIVTAAAEVS 697  
QY 740 KEFATTEKTESKIAATTTQVST-----TTQDTPPKIITLTKTTLA-PKVTITTKT 792  
Db 698 DIAIPVDPVPOEIAVAEIPETDTPKAEVIVEQSIPIEAPVPEVSKYAEVVISAPAA 757  
QY 793 ITTTEIMNKPEETAKPKDRATNSKATTPKPKQKPTKAPKPTSTKPKTPRVRKPTTPT 852  
Db 758 EVPITAGDNPDNTSGVISEVVPVIAEKPVVEVPTS--EIQSSSPS--DSVPVAKITPL 813  
QY 853 PKMTSTMPELNPTSIAEAM---LQTTTRPNOTPSKLVEN-----PKS-895  
Db 814 LRDLQTDVSLLAATLDAIGELKLDQKARNOQVMDRLCEIEKILGPPKS 864

RESULT 11  
FPI\_MYTGO STANDARD; PRT; 872 AA.  
AC Q25434;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCFFP1).  
GN FPI.  
OS Mytilus coruscus (Sea mussel).  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
OC Mytiloidea; Mytilidae; Mytilus.  
OX NCBI\_TaxID=42192;  
[1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Foot;  
RX MEDLINE=96394686; PubMed=8798340;  
RA Inoue K., Takeuchi Y., Takeyama S., Yamah E., Yamazaki F., Odo S.,  
RA Harayama S.  
RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and  
its evolutionary implications.";

Db 470 PTYKPKPSYPTTKPKITYPPTYKRRKPSYPTTKQKPSYPTTKSKSSYPTTKSKSKTYP 529  
QY 508 TT--PKGAPTTLKEPAPTTKPKAPKELAPTTTKEPTSTTSDKAPPT--TPKGTAPTTPK 564  
Db 530 PTYKPKITYPTTK-----PKSPSPSYKPKTKTTPPTYKPKIRYPTTKPKASYPPTYK 583  
QY 565 EPAPTTKPEAPTTKPGTAPTTLKEPAPTT-----KPKAPKELAPTTTKGPTSTTSDKPA 620  
Db 584 -----PKITYPTTK-PKSYPTTKPKQKPSYPTTKSKSSYPTTKSKSKTY 628  
QY 621 PTT--PKETAPTTKPEAPTTK--KPAPTTPT-----PPPTTSEVSTPTTTTKEPTT 669  
Db 629 PPTYKPKITYPTTK-PKPSYPTTKPKIRYPTTKPKKSYPTTKPKKSYPTTKPKK 687  
QY 670 IHKSDESTPELSAETPKALENSPEKGVPTTKPAATKPEMTTAKDKT-----TERDL 725  
Db 688 Y---PPSYKPKKTYPTTK-----PKISYPTTK-----TKPSYPAKYRKTSYPTTKPK 735  
QY 726 RTTPTTTAAKMTKETATTTTEKTTESKITATTTQVSTTTTQDTTPFKITTKTTLAPK 785  
Db 736 ISYPTTKAKP-----SYPTTKPKPSYVSSYKPKIRYPTTKPKPSYVSSYK-----PK 785  
QY 786 VTTTKTTTTEIMNKPEETAKPKDRATNKAATPKPQ--KPKAPKPK--TSKKKPK--T 840  
Db 786 I-----RYPTTKPKP-----SYASSYKPKIRYPTTKPKPSYVSSYKPKIT 827  
QY 841 MPRVRKPKTT--PTPRKMTSTMPELNPTSRIAEMLQTTTRPNQ 882  
Db 828 YPTTKPKISYPTTKPKITYPTTKPKKISYPTTKPKISYPTTKPKISYPSQ 871

RESULT 12  
MUC1\_XENLA STANDARD; PRT; 662 AA.  
AC Q05049;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE INTEGRATORY MUCIN C.1 (FIM-C.1) (FRAGMENT).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipelidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_Taxid=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RX MEDLINE=93077556; PubMed=1447205;  
RA Hauser F., Hoffmann W.;  
RT "p-domains as shuffled cysteine-rich modules in integumentary mucin  
polymorphism.";  
RL J. Biol. Chem. 267:24620-24624(1992).  
CC -I- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL  
INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.  
CC -I- SUBCELLULAR LOCATION: SECRETED.  
CC -I- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN  
MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
CC -I- TISSUE SPECIFICITY: SKIN.  
CC -I- PTM: EXTENSIVELY O-GLYCOSYLATED.  
CC -I- SIMILARITY: CONTAINS 6 P-TYPE (TREFOL) DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L02115; AAA74725.1; .  
CC DR PIR; A45155; A45155.  
CC DR HSP; P04002; LWFA.

DR InterPro; IPR000519; P\_trefoil.  
DR Pfam; PF00088; trefoil; 6.  
DR SMART; SM00018; P; 6.  
DR PROSITE; PS00025; P\_TREFOL; 6.  
KW Repeat; Amphibian skin; Glycoprotein; Alternative splicing.  
FT NON\_TER 1  
FT DOMAIN 81 144 8 X 8 AA APPROXIMATE TANDEM REPEATS,  
FT REPEAT 81 88 ALA/THR-RICH.  
FT REPEAT 89 96 1-1.  
FT REPEAT 97 104 1-2.  
FT REPEAT 105 112 1-3.  
FT REPEAT 113 120 1-4.  
FT REPEAT 121 128 1-5.  
FT REPEAT 129 136 1-6.  
FT REPEAT 137 144 1-7.  
FT DOMAIN 161 201 1-8.  
FT DOMAIN 218 301 P-TYPE 1.  
FT REPEAT 218 224 8 X APPROXIMATE TANDEM REPEATS, THR-RICH.  
FT REPEAT 225 239 2-1.  
FT REPEAT 240 249 2-2.  
FT REPEAT 250 259 2-3.  
FT REPEAT 260 275 2-4.  
FT REPEAT 276 287 2-5.  
FT REPEAT 288 294 2-6.  
FT REPEAT 295 301 2-7.  
FT DOMAIN 306 347 2-8.  
FT DOMAIN 353 394 P-TYPE 2.  
FT DOMAIN 402 522 P-TYPE 3.  
FT REPEAT 402 411 12 X APPROXIMATE TANDEM REPEATS,  
FT REPEAT 412 419 THR-RICH.  
FT REPEAT 432 443 3-1.  
FT REPEAT 444 453 3-2.  
FT REPEAT 454 460 3-3.  
FT REPEAT 461 472 3-4.  
FT REPEAT 473 479 3-5.  
FT REPEAT 480 491 3-6.  
FT REPEAT 492 498 3-7.  
FT REPEAT 499 515 3-8.  
FT REPEAT 516 522 3-9.  
FT DOMAIN 525 566 3-10.  
FT DOMAIN 572 613 3-11.  
FT DOMAIN 620 661 P-TYPE 4.  
FT DISULFID 162 188 P-TYPE 5.  
FT DISULFID 172 187 P-TYPE 6.  
FT DISULFID 182 199 BY SIMILARITY.  
FT DISULFID 307 333 BY SIMILARITY.  
FT DISULFID 317 332 BY SIMILARITY.  
FT DISULFID 327 344 BY SIMILARITY.  
FT DISULFID 354 380 BY SIMILARITY.  
FT DISULFID 364 379 BY SIMILARITY.  
FT DISULFID 374 391 BY SIMILARITY.  
FT DISULFID 526 552 BY SIMILARITY.  
FT DISULFID 536 551 BY SIMILARITY.  
FT DISULFID 546 563 BY SIMILARITY.  
FT DISULFID 573 599 BY SIMILARITY.  
FT DISULFID 593 598 BY SIMILARITY.  
FT DISULFID 621 647 BY SIMILARITY.  
FT DISULFID 631 646 BY SIMILARITY.  
FT DISULFID 641 658 BY SIMILARITY.  
FT VARIANT 276 276 K -> E.  
FT VARIANT 354 354 K -> R.  
FT VARIANT 415 415 T -> A.  
SQ SEQUENCE 662 AA; 67774 MW; F085277F1ED2FD40 CRC64;

Query Match 9.9%; Score 509.5; DB 1; Length 662;  
Best Local Similarity 28.1%; Pred. No. 4.9e-14;  
Matches 223; Conservative 48; Mismatches 276; Indels 247; Gaps 30;

QY 92 TTNKQSTDGKETTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEP 151  
 Db 3 TTAATAAATGKDTTAAAGSAAAEKTA--AGEVSAPPT--AAVAANGEDATT----- 51  
 QY 152 TPTTPKEPAPTTPKEPTTTIKSAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPK 211  
 Db 52 -----AAATAAEETAAAGEAPTTPAPTAAAGKAPTTPAAAGAPTTPAT 103  
 QY 212 EPAPTTPKEPAPTTPKTS-APTTPKEPAPTTPKBPAPTTPKAPTTPKAPTTPK 269  
 Db 104 GKAPATAAAPVPTTAAAGAPTTPAAAGAPTTPAAAGAPTTPAAAGAPTTP 163  
 QY 270 --PTTPKEPAPT--TPKEPAPTAPKBPAPTTPK-----EPAPTTPKEPAPTTPKE 314  
 Db 164 VKPSKREMGSGKLTKKO-----CKKNCCFPKGGHGHCFHFKPKGSHSEHTTTTK- 218  
 QY 315 PSPTTPKEPAPTTPKTSAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 374  
 Db 219 -----APTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 245  
 QY 375 KKPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 434  
 Db 246 -----TTTTKATPTT-----TTTTKATPTT-----TTTTKATPTT----- 270  
 QY 435 PTTPEKAPTTPPELAPTTPPEPTTPPEPAPTTPKAAAPTTPKAPTTPKAPTTP 494  
 Db 271 -----TTTTKATPTTPTTTP-----TTTTKATPTTPTTTP----- 289  
 QY 495 KEAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 554  
 Db 290 -----TTTTKATPTTPTTTPS-----TTTTKATPTTPTTTPS----- 328  
 QY 555 PKG-----TAPTTP-----PKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 604  
 Db 329 TKGCCFDSISIQKWCFTLSOVADCKVEPQVDCGRGIT-----ADQCRKNCCFDS 384  
 QY 605 PTTTKGPTSTSDKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 664  
 Db 385 ISGTWKCFYSTQVAA--TKTTTTPTTTTTTTTTTTTTTTTTTTTTTTTTTTT 436  
 QY 665 KEPTTIHKSFDSTSELSNAEPKALENSPKPGVPTTKTAAATKPEMTTAKDKTTERD 724  
 Db 437 TTTTTT-----TTKA-----TTTTTTTTTTTTTTTTTTTTTTTTTTT 463  
 QY 725 LRTTPTTTTAAKMKKATTTTETKTESKITATTTQVSTTTTQDTPPKITTLKTTTAP 784  
 Db 464 -TTTTTTTTT-----TTTTTKAT-----TTTTTTTTTTTTTTTTTTTTTTTTT 510  
 QY 785 KVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKQKPKKPKKSTTKPKK---T 840  
 Db 511 TTTTTKATTTTTTSGECKME-----PSKRADCGVPGITESOCRSKGCDFSSIPQTKWCFYS 566  
 QY 841 MPRVRKPKTTPR 854  
 Db 567 LQVADCKVAPSSR 580

RESULT 13  
 ID RPB1\_HUMAN STANDARD; PRT; 1970 AA.  
 AC P24928;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).  
 GN POLR2A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92178992; PubMed=1542581;

RA Wintzerith M., Aker J., Vicaire S., Vigneron M., Keding C.;  
 RT "Complete sequence of the human RNA polymerase II largest subunit";  
 RL Nucleic Acids Res. 20:910-910(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95347616; PubMed=7622068;  
 RA Mita K., Tsuji H., Morimyo M., Takahashi E., Neno M.,  
 RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;  
 RT "The human gene encoding the largest subunit of RNA polymerase II";  
 RL Gene 159:285-286(1995).  
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC -!- THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; X63564; CAA5125.1; .  
 CC EMBL; X74874; CAA52862.1; .  
 CC EMBL; X74873; CAA52862.1; JOINED.  
 CC EMBL; X74872; CAA52862.1; JOINED.  
 CC EMBL; X74871; CAA52862.1; JOINED.  
 CC EMBL; X74870; CAA52862.1; JOINED.  
 CC PIR; S21054; S21054.  
 CC MIM; 180660; .  
 CC InterPro; IPR000684; RNA\_polII\_repeat.  
 CC InterPro; IPR000722; RNA\_pol\_A.  
 CC InterPro; IPR002879; RNA\_pol\_A2.  
 CC Pfam; PF01854; RNA\_pol\_A; 1.  
 CC Pfam; PF00623; RNA\_pol\_A2; 1.  
 CC PROSITE; PS00115; RNA\_POL\_II\_REPEAT; 43.  
 CC Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 CC DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 CC ZN\_FING 71 87  
 CC C2H2-TYPE (POTENTIAL).  
 CC DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC FT CONFLICT 1067 1067 W -> L (IN REF. 2).  
 CC FT CONFLICT 1449 1449 D -> Y (IN REF. 2).  
 CC SQ SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;

Query Match 9.8%; Score 503.5; DB 1; Length 1970;  
 Best Local Similarity 33.3%; Pred No. 2e-13;  
 Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;  
 QY 151 PTPPTPKPAPTTPKE-PTPTTIKSAPTTPKEPAPTTPKAPTTPKAPTTPKE 204  
 Db 1507 PSPMGISPMPTPNQGATPAYGAWSPVSGMTPCAAGFSPSAASDASGSPGSPAWS 1566  
 QY 205 PAPTTPKEPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKEP-APTTPKEPT 262  
 Db 1567 PTPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSP 1624  
 QY 263 TTPKEPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKEPTTPKE 322  
 Db 1625 TSPSY-SPTSPNVSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1671  
 QY 323 PAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 382





CC EMBL; M37227; AAA41693.1; ALT\_FRAME.  
DR EMBL; X13804; CAA32038.1; ALT\_FRAME.  
DR EMBL; M21964; AAA41695.1;  
DR EMBL; J04517; AAA41692.1;  
DR PIR; A30796; A30796.  
DR PIR; A25649; A25649.  
DR PIR; B25649; B25649.  
DR PIR; S02003; S02003.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.  
FT DOMAIN 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.  
FT CONFLICT 164 164 L -> I (IN REF. 2).  
FT CONFLICT 185 185 I -> S (IN REF. 2).  
FT CONFLICT 193 193 L -> T (IN REF. 2).  
FT CONFLICT 199 199 M -> T (IN REF. 2).  
FT CONFLICT 346 346 K -> N (IN REF. 1).  
FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).  
FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).  
FT CONFLICT 485 485 P -> S (IN REF. 2).  
FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).  
FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).  
FT CONFLICT 727 727 A -> V (IN REF. 4).  
FT CONFLICT 759 759 AAP -> GST (IN REF. 4).  
FT CONFLICT 769 769 T -> L (IN REF. 2).  
FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).  
SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;  
Query Match 9.6%; Score 497; DB 1; Length 831;  
Best Local Similarity 27.6%; Pred. No. 1.8e-13;  
Matches 195; Conservative 59; Mismatches 307; Indels 146; Gaps 35;  
QY 37 POTSITQHNKVTSPKIIITAKIPNPSPNSDTSKETSITVKNKETTIVETKETTNNK 96  
DB 197 PSMST--HIKVKSEKIKVVE-----KSEKEIVIVEEQTEEQIQTVEETED 242  
QY 97 TSDGKEKTSKETSIAKDLAPTSKVLAKPTPKAETTTGKPAITTTKEP---T 152  
DB 243 KEAQGESEEEAEGBGEAAATTS-----PPAEAAASP-----ERETKSPVKEAKSPA 293  
QY 153 PTPPKKPA-STTPKEPTPTTIKSAPTTPKPA-----PTTTKSAPTTPKPA-----PTTTK 203  
DB 294 PAEAKSPAESAEPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKS 352  
QY 204 EP-----APTTPKPA-----PTTTKPA-----PTTTKSAPTTPKPA-----PTTPK 245  
DB 353 SPGEAKSPAESAEPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPV 411  
QY 246 --PTTPKPA-----PTTPKPA-----TPTTPKEPATTPKPA-----APTTPKPA-----PTA 287  
DB 412 KSPVEAKSPAESAEPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPV 470  
QY 288 PKPAPTTPKPAAPTTPKPAAPTTPKPAAPTTPKPAAPTTPKPAAPTTPKPAAPTTP 338  
DB 471 AKSPAEP--KSPAGAKPPAEAKSPAESAEPVKSAPVKSAPVKSAPVKSAPVKSAPV 528  
QY 339 APTTTKS-----APTTPKEPSPTTTKPA-----PTTPKPAAPTTPKPAAPTTP 389  
DB 529 VKEGAKSLAEAKSPAESAEPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPV 588  
QY 390 PKPAPTTPKPAAPTTPKPAAPTTPKPAAPTTPKPAAPTTPKPAAPTTPKPAAPTTP 449  
DB 589 AKPPAEAKSPAESAEPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAPVKSAP 644  
QY 450 APTTPEPTPTTPKEAPTTPKAAAPTTPKPAAPTTPKPAAPTTPKPAAPTTPKPAAPT 509  
DB 645 ---KAKEPPKVEEKTPTPKTEVKESKDEAPKQKP--KAEKEKPLETKPKDSGEA 700  
QY 510 PKGTAPTTLKEPATTPTKKPAKPAKELAPTTTKEPTSTTSIDKPAPTTPKGTAPTTP 569

RESULT 15  
NFH\_RAT  
ID NFH\_RAT STANDARD; PRT; 831 AA.  
AC P16884; Q63368;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)  
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).  
GN NEFH OR NFH.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=89065087; PubMed=3143606;  
RA Breen K.C., Robinson P.A., Wion D., Anderton B.H.;  
RT "partial sequence of the rat heavy neurofilament polypeptide (NF-H).  
RT Identification of putative phosphorylation sites.";  
RL FEBS Lett. 241:213-218(1988).  
RN [2]  
RP SEQUENCE OF 37-831 FROM N.A.  
RX MEDLINE=88309090; PubMed=2457365;  
RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,  
RA Jolles P.;  
RT "the large neurofilament subunit (NF-H) of the rat: cDNA cloning and  
RT in situ detection.";  
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).  
RN [3]  
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.  
RX MEDLINE=87080760; PubMed=2878828;  
RA Robinson P.A., Wion D., Anderton B.H.;  
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide  
RT (NF-H).";  
RL FEBS Lett. 209:203-205(1986).  
RN [4]  
RP SEQUENCE OF 318-831 FROM N.A.  
RX MEDLINE=89184647; PubMed=2928342;  
RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,  
RA Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;  
RT "Cloning of a cDNA encoding the rat high molecular weight  
RT neurofilament peptide (NF-H): developmental and tissue expression in  
RT the rat, and mapping of its human homologue to chromosomes 1 and  
RT 22.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).  
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783  
CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
Db 701 KKEEA-----KEKKAAPPEETPAKLGVKEEAKPKEKAEDAKA-----KEPSKP 744
Qy 570 TPKEPAPTTPKGTAPTTILKEPAPTTPKKPAKPAKELAPTTTKGPTSTTSDDKPAPTPKETAP 629
Db 745 SEKE-----KPK-----KEVPAAPEKKDTKE-----ERTTESKKREKPKMEAK 784
Qy 630 TTPKEPAPTTPKKPAPTTPEPPPTTSEVSTPTTTTKEPTTIHKSPDE 676
Db 785 AKED-----KGLPQEPSKPKTEKAEKSSSTDQKDSQPSEKAPED 824
```

Search completed: April 26, 2002, 16:19:35  
Job time: 360 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:52 ; Search time 49.78 Seconds  
(without alignments)  
437.589 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1167  
Perfect score: 5155  
Sequence: 1 VKDNKNRTRKKRTPKPPV.....GKPDGLTLRLNGLVAFRG 968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522	10.1	805	4	US-09-103-429A-4
2	506.5	9.8	744	6	5202236-25
3	504	9.8	1837	3	US-08-928-361B-5
4	498.5	9.7	786	4	US-09-103-429A-3
5	489	9.5	1721	3	US-08-700-651-5
6	489	9.5	1721	3	US-08-928-361B-6
7	488.5	9.5	826	1	US-07-638-431-2
8	488.5	9.5	826	5	PCT-US92-00018-2
9	476.5	9.2	652	6	5202236-13
10	452	8.8	960	4	US-09-219-849-5
11	423.5	8.2	1867	2	US-08-479-537A-5
12	423.5	8.2	1867	4	US-09-083-116-5
13	423.5	8.2	2035	2	US-08-479-537A-2
14	423.5	8.2	2035	4	US-08-479-537A-2
15	419.5	8.1	2476	2	US-09-083-116-2
16	419	8.1	1185	4	US-08-276-967-2
17	417	8.1	829	1	US-08-642-255-132
18	417	8.1	829	1	US-08-397-633A-53
19	417	8.1	837	1	US-08-175-155-68
20	417	8.1	837	1	US-08-477-509B-103
21	417	8.1	837	1	US-08-642-255-101
22	417	8.1	837	2	US-08-707-237A-75
23	417	8.1	837	3	US-08-482-085B-103
24	417	8.1	897	1	US-08-397-633A-50
25	413.5	8.0	907	3	US-08-783-774-2
26	413.5	8.0	907	5	PCT-US95-04611A-19
27	404.5	7.8	408	1	US-07-609-716-65

28	404.5	7.8	408	4	US-08-475-411A-65	Sequence 65, Appl
29	404.5	7.8	408	4	US-08-478-029A-65	Sequence 65, Appl
30	398.5	7.7	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	7.7	682	1	US-08-397-633A-36	Sequence 36, Appl
32	381	7.4	1848	4	US-08-296-791-6	Sequence 6, Appl
33	381	7.4	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
34	376	7.3	960	4	US-09-219-849-6	Sequence 2, Appl
35	374.5	7.3	1537	1	US-08-325-267A-2	Sequence 4, Appl
36	371.5	7.2	1231	3	US-08-904-263A-4	Sequence 114, Appl
37	371	7.2	761	2	US-08-707-237A-84	Sequence 2, Appl
38	371	7.2	762	1	US-08-642-255-114	Sequence 84, Appl
39	371	7.2	762	1	US-08-397-633A-26	Sequence 114, Appl
40	370.5	7.2	1084	1	US-08-642-255-62	Sequence 26, Appl
41	368	7.1	762	1	US-08-642-255-120	Sequence 120, App
42	368	7.1	762	1	US-08-397-633A-31	Sequence 31, Appl
43	367.5	7.1	1187	1	US-08-320-559-28	Sequence 28, Appl
44	367.5	7.1	1187	3	US-08-545-860D-28	Sequence 28, Appl
45	367.5	7.1	1187	5	PCT-US94-04496-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1  
US-09-103-429A-4  
; Sequence 4, Application US/09103429A  
; Patent No. 6187558  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 No. 6187558th Tioga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/103.429A  
FILING DATE: 24-JUN-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 805 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-103-429A-4

Query Match 10.1%; Score 522; DB 4; Length 805;  
Best Local Similarity 27.8%; Pred. No. 5.2e-26;

FILING DATE: 24-NOV-1986  
APPLICATION NUMBER: 650,128  
FILING DATE: 13-SEP-1984  
SEQ ID NO: 25  
LENGTH: 744  
5202236-25

Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;

QY 125 TSVKLAKPTPKAETTTKGPALITPKPTTTPKEPASTTPKPTTTPKTSAB-----TTP 179  
Db 9 TALGLVAARPEVSDAEKNFALHEPHDPX-PAEQXLLLPXEYDCTKFYYCEYGLKFIAP 66  
QY 180 KEPAPTT-----TKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 227  
Db 67 RDCAPGTEFKESAQTCVHAALACGLPGPAET---QAPATQ---APTTQ---APTTT 119  
QY 228 SAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKPTTTPKEPAPTTTPKEPAPTA 287  
Db 120 QAPTTTQAATTT-----QAPTTQAATTT-----QATTTQAATTT----- 156  
QY 288 PKPAPTTTPKEPAPTTTPKEPAPTTTPKAPTTTPKAPTTTPKAPTTTPKAP 347  
Db 157 ---QAPTTTQ---APTTTQ---APTTQ---APTTQ---APTTQ---APTTTQA 198  
QY 348 TTPKPSPTTTPKEPAPTTTP---KEPAPTTTPKAPTTTPKEPAPTTTPK----- 391  
Db 199 TTPAATTTPAAT---TPAATTTPAATTTPAATTTPGVPAPTS-----APVMPICELLPGCADFD 253  
QY 392 -----EPAPTTTPK 399  
Db 254 IHLPHDKYCNLFYQCSNGYTFEQRCEGLYFNPIYVORCDSPANVECDGELSPAPVIE 313  
QY 400 K-----PA----- 402  
Db 314 GNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERRCGAGTHFSFELQ 373  
QY 403 -----PTAP---KEPAPT-----TPK 417  
Db 374 QCDHIELWGCTLPGGSEBEVDVEDACTGWVCTPEIWEPLNGCPADFSIDHLLPHES 433  
QY 418 -----APTPKKTPTT-----PEKLAPTTPKEPAP 443  
Db 434 DCGQYLCQVHGQTARPCGNLHSPATQSCSPVTAGCQVFECDSNQCTSTAAPTAAP 493  
QY 444 TPEELAPTTPEPTTPKEPAPTTTPKAAAPNTPKBPAPTTTPKEPAPTTTPKEPAPTTPK 503  
Db 494 TAAPTAAPTAAPTAAPSVVPPA---TPPATAAPVPPPTAIP---PAPTAAPTAAPTAA 548  
QY 504 ETAPTT---PKGAPTTLKEPAPTTTPKPKAPKELAPTTTKEPTSTSDKAPTTTPKGATP 560  
Db 549 PESPTTVVPPTAAPTA---APIT---AVPEIPTVTSAPTAAPT---AAPTAAPTAAPT 598  
QY 561 TTPKEPAPTTTPKEPAPTTTPKGATPTLKEPAPTTTPKPKAPKELAPTTTKEPTSTSDKAP 619  
Db 599 TTAAPVEIPTT-----VTSPTTAAPT---AAPANT-----TVPPTAAPTAAPT 641  
QY 620 APPTPKETAPTTTPKEPAPTTTPKPKAPPTTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTP 679  
Db 642 AP-----NTTVVAPPTAAPT---AAPANTTVVPPPTAAPTAA---PTVAH-----AP 685  
QY 680 ELSAEPTRKALENSPKPGVPTTKTPAATKPE 711  
Db 686 NTAAP-----VTTTSAPATT-PE 703

RESULT 2  
5202236-25  
; Patent No. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
; PROTEIN  
; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,762  
; FILING DATE: 25-MAY-1990  
; APPLICATION NUMBER: 82,456  
; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945

Query Match 9.8%; Score 506.5; DB 6; Length 744;  
Best Local Similarity 29.9%; Pred. No. 4.6e-25;  
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 51 PKITTAKPINPRSPNPSDTSKETSITVNKETTVEKETTITNNKQSTDCGKETTSAKE 110  
Db 24 PKMTVPPTPKPSYPP-----IVKSKPTY-----PKKIT----- 53  
QY 111 TOSIEKTSAKDLAPTSKVLAKPTKPAETTTKGPALITPKPTTTPKEPTTPKPE-----PASTTPKE 166  
Db 54 -----YPPTYK---AKPS-----YPPTYKPKKTYPPPTPKKLYPPPTPKKLYPPPTPKP 92  
QY 167 PTPTTIKSAPT---TPKEPAPTTTKSAPTTTPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPA--P 223  
Db 93 SYPTPKSKPTTKKILTYPPYKAKPSYPPPTPKPKTYPPPT---YKPKLYPPPTPKKASYP 151  
QY 224 TTTKSAPTTTPKEPAPTTTPKAPTT---TPKEPAPTT-----PKPTTTPTPKEPAPTTTPKEP-- 275  
Db 152 PTYKPKPSYPP---PSYTKTKTYPPPTPKKLYPPPTPKPSYPPSYKPKKTYPPPTPKKLT 209  
QY 276 -APTTTPKEPA---PTAPKKA---PTTPKEPA---PTTPKEPA---PTTPKEPS---PTTPKEPA 324  
Db 210 YPTTYKAKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPS 269  
QY 325 -PTTPKSAPTTTPKEPAPTTTKSAPTTTPKPSPTTTPKEPA---PTTPKEPA---PTTPKPKP 377  
Db 270 YPTTYKAKPSYPP---YPTTYKAKPSYPP---PTYKAKPSYPPPTPKPSYPPPTPKPSYPPPTPK 322  
QY 378 A---PTTPKEPA---PTTPKEPA---PTTPKKA---PTAPKEPAPTTTPKEPAPTTTPKLTPTT 429  
Db 323 SYPTTYKAKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPS 382  
QY 430 PEKLAPTTPKE---KPAPTTPPELAPTTTPKEPT---PTTPKEPA---PTTPKEPA---AAPNT---P 478  
Db 383 TYKAKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPP 438  
QY 479 KEPAPTTTPKEPAPTT-----PKAPTTTPKEPAPTT---PKGAPTTLKEPAPTTTPKPKAPK 532  
Db 439 SYPTTYKPKISYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPS 498  
QY 533 ELAPTTTKEPTSTSDKAPTTTPKGATPTPKEPAPTTTPKEPAPTTTPKGATPPTLKEPA- 591  
Db 499 SYPTTYKAKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPS 546  
QY 592 -PTTPKPKAPKELAPTTTKGPT---STSDKAPTTTPKEPAPTTTPKEPAPTTTPKPA---PT 646  
Db 547 PPTTYKAKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPS 601  
QY 647 ---TPETPPPTTSEVSTPTTKE---PTTIHKSPD-----ESTPELSAEP---PKALENS 693  
Db 602 YKAKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPP 661  
QY 694 PKPGVPTTKTPAA-----TKPEMTTTAKDKTTERDLRTTPPTTAAAP-----KMT 739  
Db 662 PSYP---PTYKAKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPP 719  
QY 740 KETATTTTEKTESKITATTT 759  
Db 720 YKAKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPPPTPKPSYPP 739

RESULT 3  
US-08-928-361B-5  
; Sequence 5, Application US/08928361B  
; Patent No. 6071518

GENERAL INFORMATION:  
APPLICANT: Petersen, Carolyn  
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
TITLE OF INVENTION: SPECIES INFECTIONS  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-1840  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928.361B  
FILING DATE: 12-SEP-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/026.062  
FILING DATE: 13-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: VERNY, Hana  
REGISTRATION NUMBER: 30,518  
REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
TELEPHONE: 650-324-1677  
TELEFAX: 650-324-1678  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1837 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-928-361B-5

Query Match 9.8%; Score 504; DB 3; Length 1837;  
Best Local Similarity 22.3%; Pred. No. 1.8e-24;  
Matches 279; Conservative 81; Mismatches 481; Indels 412; Gaps 40;  
QY 55 TAKDINPRPSLPNSDTSKETSIVNKET-----TVETKETTNNKQST-----DGK- 102  
Db 215 TGLPTDYPNCPNPVTVGNLVSRSSTGKTIPNTYAGVYRSNETKTTPSANTNELLVDPKI 274  
QY 103 ----EKTSKAKTSQIEKTSKDLAPTISKVLAKPTPKAETTTKGPALTTTPKEPTTPPK 158  
Db 275 NACNSENSEFQVOIFDMGSKVYIPYTKCVGVKHTTTTTT-----TTTTTTTTTTTT 329  
QY 159 PASTTPKEPTPTTKSAPTTKPEPAPTTPKSAPTTPKEPAPTTPKPEPAPTTPKPEPAPT 218  
Db 330 TTTTNTTTTTTTTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 383  
QY 219 KEPAPTTKSAPTTKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPT 278  
Db 384 TTTTNTTTTTTTTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 443  
QY 279 TPKEPAPTAKKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPE 338  
Db 444 TTTTNTTTTTTKKPTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 497  
QY 339 APPTTKSAPTTKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPK 376  
Db 498 TTTTNTTTTTTKKPTTTTATTTTSETSVIKPDENCWLEKNGCEAKGATVGVIG 557  
QY 377 ----- 376

Db 558 KQRIENGMAFTMIPNDTHVRFKVKVDGNTISVRCGKGAGKLEFPDRSLDFTIPPVA 617  
QY 377 -----PAPTTPKE-----PAPTTPKE-----PAPTTPKE-----PAPTTPKE 388  
Db 618 GHNSCSIIIVSGGGKTHVSPYSGKDVLSAPIQCELFENEVYCDTCTAKYGAHSGYQ 677  
QY 389 TPKEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPE 448  
Db 678 TSADFVTTTAKPTTTT--TGAPGQPTTTTSGSPKPTTTTAKATTTT-----TILNP 729  
QY 449 LAPTPPEPTTPPEPAPTTPKAAAPNTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPE 508  
Db 730 IITTTQKPTTTT-----TT--KVPGRPIATTTTTLKPIVTTTTRKATTTTTPVT 780  
QY 509 TPKEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPE 549  
Db 781 T-----TTTTKRDGMTTTTPLDIDGIEITPIPIEKMLDKYRMIDYNSGLLLDSNDE 835  
QY 550 PAPTTPKG-TAPTTPKEPA-----PPTPKPEPAPTTPK-----PPTPKPEPAPTTPK 580  
Db 836 PIPGSAQGIADTSNLFPPVQTHKSTGLPIDPMVGLPDPKSGNLVHPYTNQTMGSLSVY 895  
QY 581 -----GTAPTTLK-----EPAPTTPKPEPAPKELAPTTT----- 608  
Db 896 LAAKNLTVDDETYGLPIDTLGYPIDVSLIPFPNETGELFDPIDSEIMNGTIAGIVSG 955  
QY 609 -KGPTSTTSKPEPAPTTPKE-----TAPTTPKEPAPTTPKPEPAPTTPKPEPAPTTP 651  
Db 956 ISASESLLSQKSAPIDPATNMVVGEGGLNPATGMIPGSLGSEQTPFSEIEDGGII 1015  
QY 652 PPTTSEVSTPTTKEPTTIHKSPDESTE-----LSAEPKPK 688  
Db 1016 PP---EVAANADKFKLSIPSPVPSIPEKQKIDISSELMYDIESGRUGVSKRPIG 1072  
QY 689 ALENSKPEGVPTTKTPAATKPEMTTAKDKTTERDLRTPTTETTAAPKMTKETA---T 744  
Db 1073 SIAGDLN-----PIMKTPTQT-----DSVTGKPI-----DPTTGLP-FNPTGHLNP 1114  
QY 745 TTEKTTESKITATTTQVSTTTQDTPPKITTLTKTTTTLAPK--VTTTKKTTTITIMNKP 802  
Db 1115 TNNNTMDSFAGAYKAVSNGIKTQNVYGLPVDITGL-PKDPVSDIPFNSNTGELVD-- 1171  
QY 803 EETAKPKDRAT-----NSKATTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPT 846  
Db 1172 PSTGKPINNYTAGIVSGKRLGPPIDENGNLFPDSTKPLPIDGNOLVNPENSTSVSGSTS 1231  
QY 847 PKTTTPPKMNS---TMPE-----LNPTSRIAEAMLOTTTTPRNPOTPSKL 888  
Db 1232 GSTKPKGIPVNGGVVPEDEAKDQADKDGKGLIIVPPTNSINKDPVTNTQYSNTTGN-- 1289  
QY 889 VEVNPKSEDAGGAGCETPHMLLRPHVFMPEVTPD-----MDYLPVPVNGQIINP-- 938  
Db 1290 --INF--ETGKVIPLGSLGSLNYSFNTQQQDEITGAPVDVTGGLPVDPSGTGEIIPAT 1345  
QY 939 -----MLSDETNICNGK-----PVDGLTTLRNGTLV 964  
Db 1346 KLPIPGSVAGDEILTEVLNITTTDEVTGLPIDLETGLPRDPVSGLPQLPNGTLV 1398

RESULT 4  
US-09-103-429A-3  
; Sequence 3, Application US/09103429A  
; Patent No. 6187558  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 No. 6187558th Tloga  
; CITY: Ithaca

STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/103,429A  
FILING DATE: 24-JUN-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 786 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-103-429A-3

Query Match 9.7%; Score 498.5; DB 4; Length 786;  
Best Local Similarity 29.1%; Pred. No. 1.6e-24;  
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

QY 132 PTPKAEETTKGALTTKPEPTTPKPEASTTPKPEPTTIKSAPTTPKPEAPTITKSAP 191  
DB 93 PGPPAE-TTQAPAT- - - - -QAPTITQAPTIT- - - - -QAPTITQ- - - - -ATTITQAP 135  
QY 192 TTPKPEAPTTPKPEAPTTPKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKE 251  
DB 136 TTTQ-ATTIQ-ATTI- - - - -QAPTITQ-APTITQAPTITQ-APTITQ-APTITQ- 183  
QY 252 PAPTTPKEPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEA- 308  
DB 184 -APTITQAPTITQAAATTTPAATTTPAATTTPAATTTPAATTTPGVPAPTSAPVWPPIC 241  
QY 309 - - - - -PTTKPEPSPTPK- 321  
DB 242 ELLPNCGPADFIIHLPHDKVCNLFYOCNNGYFEQRCPEGLYENPVVQRCDSAPNVEC 301  
QY 322 - - - - -EPAPTTT- - - - -PTTKPEPSPTPK- 321  
DB 302 DGEISAPVPVTEGNEDEDIDGLDNGCPANFEIDLPHGNRCDDKYQCVHGNLVERR 361  
QY 329 - - - - -KSAPTTKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 361  
DB 362 CGAGTHFSFELQOCDHIELVGTCLFGGESEVDVDEDA-CTGWYCPTEPIEWEPLNCP 420  
QY 362 APTT- - - - -KPEP- - - - -APTTPKPEAPT- - - - -APTTPKPEAPT- 381  
DB 421 ADFSIDLPHESDCQYLQCVHGQTIARPCPNLHFSFATQSCSPVTAGQCVFECDS 480  
QY 382 - - - - -KPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 438  
DB 481 NOCTSTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 533  
QY 439 EKAPPTPEELAPPTPEPTTPPEAPTTPKAAAPNT- - - - -KPEAPTTPKPEAPTTPKE 496  
DB 534 - - - - -PAPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 587

QY 497 PAPTTPKETAPTTPKGTAPTTLKBPAPTPPKKPAKELAPTTTKPTSTTSKDPAPTPPK 556  
DB 588 AAPTAAPTAAAPTAVPEIPTVTTSPTTAAPTAA- - - - -APNTT- - - - -VTVP 632  
QY 557 GTAPTTPKPEAPTTPKPEAPTTPKGTAPTTLKBPAPTPPKKPAKELAPTTTKGPTSTTS 616  
DB 633 TAAPT- - - - -AAPANTVTVPPTAAPTAAAPTVAH- - - - -APNTTAAAPTSTTS 677  
QY 617 DKPAPTTPKETAPTTPKPEAPTTPPKP 643  
DB 678 - - - - -APATTPEDDIDP- - - - -PLNDPDPNP 699

RESULT 5  
US-08-700-651-5  
; Sequence 5, Application US/08700651B  
; Patent No. 6015882  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN  
; APPLICANT: LEECH, JAMES  
; APPLICANT: NELSON, RICHARD, C.  
; APPLICANT: GUT, JIRI  
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
; FILE REFERENCE: 480.19-4(HV)  
; CURRENT APPLICATION NUMBER: US/08/700,651B  
; CURRENT FILING DATE: 1997-08-14  
; EARLIER APPLICATION NUMBER: 08/415,751  
; EARLIER FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1721  
; TYPE: PRT  
; ORGANISM: Cryptosporidium parvum  
US-08-700-651-5

Query Match 9.5%; Score 489; DB 3; Length 1721;  
Best Local Similarity 22.1%; Pred. No. 1.5e-23;  
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

QY 5 KKNRTKKKTPKPPVDEAGSLDNGDFKVTTPDSTTOHKNVSTSPKITTAKPINRPS 64  
DB 116 RSNETK- - - - -TTEPSANTYAGVRSN-ETKITEPSANT- - - - -NLLVDPKI- - - - -N 158  
QY 65 LPNDSSTSK- - - - -TSLTVNKETTVETKETTNTKOTSTDCKEKTSK 109  
DB 159 APCNSENSEFQGIQFDMGSKVYIPYTKCVGKHKTITTTTTTTTTTTTTTTTT 208  
QY 110 ETOSIEKTSKADLAPTSKVLAKPTPKAETTKGALTTKPEPTTPKPEAPTTPKPEPT 169  
DB 209 - - - - -TT 240  
QY 170 TTIKSAPTTPKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKSA 229  
DB 241 TT 300  
QY 230 PTTKPEAPTTPKPPAPPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 289  
DB 301 TTTTTPKPTTT 350  
QY 290 KPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKSAPTTPKSAPTT 349  
DB 351 - - - - -TT 406  
QY 350 PKE- - - - -PSPTT- - - - -PSPTT- - - - -PSPTT- - - - -PSPTT- 357  
DB 407 TSEIESVIKPEWCKEKECEAKGATYVGVIGDKGRIENGMAFTWIPNDTHVRPFX 466  
QY 358 - - - - -TKEP- - - - -TKEP- - - - -TKEP- - - - -TKEP- - - - -TKEP 368



```

Db 467 VKDVGNTISVRCRGAGKLEFPDRSLDTFTPPVAGHNSCSIIIVSGDGKIHVSPGSKD 526
QY 369 PAPTTPKKPAPTTPKE-----PAPTTPKEPAPTTPKKPAPTAPKPEPA 410
Db 527 VSLIS-----APIQSELFNEVYDCTCTAKYGAHSGYQTSADRVTTTAKPTTTT--TGA 580
QY 411 PTPKETAPTTPKKLTPTTPEKLAPTTPKAPATTPEELAPTTPPEPTTPPEEPAPTTP 470
Db 581 PCQPTTTTSGSPSKPTTTTAKATTT-----TTLNPIITTTTQKPTTT-----TT- 627
QY 471 KAAPNTPEPAPTTPKEPAPTTPKETAPTTPKGTAPTTPKGTAPTTLKEPAPTTPKKPA 530
Db 628 --KVPCKPIATTTLTKBIVTTTAKATTTTTPVT-----TTTKKROEMTTTTL 680
QY 531 PK-----ELAPTTPKETS-----TTSKPAAPTTPKG-TAPTTPKEPAPT 570
Db 681 PDIGDIEITPIERKMLDKYTRMIDYNSGLLSDNDEPIFGSQAGQIADTSNLFVQTH 740
QY 571 PKEPAPTTPKGAPTTLKE---PAPTTPKPAKELAPTTPKGTSTTSKPAAPTTPKET 627
Db 741 KSTGLPIDPMVGLPDPKSGNLVHPYTNQWGLSVLAALKNLVTDITYG--LPIDT 798
QY 628 APTTPKEPAPTTPKKPAPTTPETPTTSEVSTPT-----TTKEPTTIHKS----- 673
Db 799 LTGYPLDPSLIPFN--PETGELEFDTSDIEMNGTIAGIVSGISASESLLSOKSALIDPA 856
QY 674 -----PDEST-----KQKIDISSELMYDIESGRIGVSKRPIPGSIAGDLNP 963
Db 857 TNNVYGEFGLLNPAATGVMIPGFLGSPSQOTQFSEIEDGGIIPPEVAANADFKLSIP- 915
QY 696 EPGVPTTKTPAATKPEMTTAKDKTTER-----DLRTPPTTTAAPKM 738
Db 916 -PSVP-----ESIP-----KQKIDISSELMYDIESGRIGVSKRPIPGSIAGDLNP 963
QY 739 TKETATITTEKTESKITATT--TQVTSSTTQDTPFKITTLTKTTTLAPKVTTTKTITTE 797
Db 964 IMKPTQDTSVTGPIDPTTGLPNPTGHLINPTNNMTDSSFAGAYKAVANGIKTON 1023
QY 798 IMNKP--EETAKPD-----RATNSKATTPKPKPTKAPKPTSTTKPKTMP----- 842
Db 1024 VIGLVPGEITGLPKDPSDIPFNSTTGELVDPTGKPIINNSTAGIVSGKPLPIEDENG 1083
QY 843 -----RVKPKTTPTPKMTS--TMPE-----862
Db 1084 NLEDPSTNLPIDGNQLVNETSTVSGSTGTTKPKPIPVNGGVVPEDEAKQADKG 1143
QY 863 -----LNPTSRIAEMLOTTTRNPQTPNSKLVENPKSDAGGAEGETPHMLLRPHVMP 917
Db 1144 KDGLIVPTNSINKDPVTNTQYSNTGNI-----INP--ETGKVPISLPGSLNYSPFNT 1197
QY 918 EVTPD-----MDVLPVPNOGIINP-----MLSDFTNICNGK----- 950
Db 1198 QQTDEITGKPDVDTGLPVDPTGGEIIDPATKLPIGSPVAGDEILTVLNTTDEVTGLP 1257
QY 951 -----PVDGLTTLRNGTILV 964
Db 1258 IDLETGLPDPVSGVLPQLPENGTLV 1281

```

```

RESULT 6
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA

```

```

; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-6

```

```

Query Match 9.58; Score 489; DB 3; Length 1721;
Best Local Similarity 22.18; Pred. No. 1.5e-23;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

```

```

QY 5 KNRTRKKTTPKPPVVDGASGLDNGDFKVTTPDTSTTQHKNVSTSPKITTAKPINRPS 64
Db 116 RSNETK---TTPEPSANTYAGVYRSN-ETKITEPSANT---NFLVDPKI-----N 158
QY 65 LPPNSDTSKE-----TSLTVNKETTVEKETTNTNKTSTGDKKETSATK 109
Db 159 APCSENSEFQGGQIFDMGSKVYIPYTKCVGKVTHTTTTTTTTTTTTTTTTTTTT 208
QY 110 ETQSIKTSADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPASTTPKEPTP 169
Db 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
QY 170 TTIKSAPTTPEPAPTTPKSAPTTPKEPAPTTPKAPATTTPKEPAPTTPKSA 229
Db 241 TTTTTPPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
QY 230 PTTTPKEPAPTTPKPPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTAPK 289
Db 301 TTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 350
QY 290 KPAPTTPKEPAPTTPKEPAPTTPKESPTTPKEPAPTTPKSAPTTPKSAPTTKSAPT 349
Db 351 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 406
QY 350 PKP-----PSPPT-----357
Db 407 TSETSVIKPDEWCWLEKNGECEAKGATYGVIGKDGRIENGMAFTMIPNDTHVRFREK 466
QY 358 -----TKEP-----APTTPKE 368
Db 467 VKDVGNTISVRCRGAGKLEFPDRSLDTFTPPVAGHNSCSIIIVSGDGKIHVSPGSKD 526
QY 369 PAPTTPKKPAPTTPKE-----PAPTTPKEPAPTTPKKPAPTAPKPEPA 410
Db 527 VSLIS-----APIQSELFNEVYDCTCTAKYGAHSGYQTSADRVTTTAKPTTTT--TGA 580

```



```

Db      607 PKKPI-----NPEESNPKPEP-----INPEDENENPLIQDEPIEPRNDSNVIP1 649
Qy      631 TPKEP-----APTPKKAPPTPEPTPEPTTSVSPTTKEPTTHKSPDESTPELSAE 684
Db      650 LPLIIPKGNINPSINPEPNSDSEVEYPRPDNENSNNTKMSKNI---PNEPIPSGDN 706
Qy      685 PTPKALENSPK-----EPGVPTTK 703
Db      707 PYKGHEERIPKPHRSNDVYVDNNVKNKKKDEFEIPNNE 745

.RESULT 9
5202236-13
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TEN; FLIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO:13
; LENGTH: 652
5202236-13

Query Match          9.2%; Score 476.5; DB 6; Length 652;
Best Local Similarity 29.7%; Pred. No. 3,3e-23;
Matches 221; Conservative 71; Mismatches 284; Indels 169; Gaps 45.

Qy      51 PKITAKPINRPSLPRNDSIDSKETSLVYNKETVETKTTTNNKOTSIDGKEKTSKAKE 110
Db      11 PMKTYPPYKKKPSYP-----YKSKPTV-----KPKIT---- 40
Qy      111 TOSIEKTSADLAPISKVLAKPTPKAETTITKGALTTTPEPTTTPKE-----PASTTPE 166
Db      41 -----YPIYK--AKPS-----YPIYKPKKITPPYKKPLTTPYTKPKP 79
Qy      167 PPTPTIKSAPV--TPKEPAPTTTKSAPTPKEPAPTTTKEPAPTTPKEPAPTTPKEA--P 223
Db      80 SYPIYKSKSPYTKPKITVPPYTKAKPSVPYTKPKKTYPT--YKDKLVPPYTKPKASYP 138
Qy      224 TTTKSAPITPKPAPATTTTKKPAV--TPKEPAPT-----PKEPTPTTPKEPAPTTPKEAPT 278
Db      139 PYYKKKPSVP--PSYKTKKTYPTTKPKLTYPTTKPKKPSYSPSYKPNKTYPTTKPKPLT 196
Qy      279 TPKEPAPAPKKPA--PTTPKEPA--PTTPKEPAPTTTKEPSPTTP-----KEPAPTTPK 329
Db      197 YP-----PYTKAKPSVPYTKAKPSVPYTKAKPSVPYTKAKPSISLIKAKPSIPTYK 252
Qy      330 SAPTTTKEPA-----PTTKSAPITKEPSPTTKEPA--PTTPKEPA-----PTTPK 376
Db      253 AKPTYKAKPTYKAKPTVSYTKAKPTVP---PTTKAKPSVPYTKAKPTYKAKPTYKAK 308
Qy      377 PA--TTTPKEPA--TTTPKEPAPTTTKEPAPAPAPAKPAPTTTKEAPITPKKLPTTPBEK 432
Db      309 PSYPTPYKAKPSVPYTKAKPSVPYTKAKPSYP---PTYKAKPSVPYTKVVKPT--YK 362
Qy      433 LAPTPPE--KPAPTTPEELAPTTPEPT--PTTPEEPA--PTTPKKAADNTPKEPAPTTPK 487
Db      363 AKPTVPSYTKAKPSVP---PTYKAKPSVPYTKAKPSVPYTK--AKPSYP---PTTKA 413
Qy      488 EPA--PTT---PKEPAPTTPKETAPT--PKGTAPTTLKEPAPTTPKPAPEKELAPTTPK 540
Db      414 KPTVSTYTKAKPSVPYTKPKKPSYPTTKAKPSYSTTKAKSSSYPTTKAKKSSYPTTKA 473

```

[illegible]

```

RESULT 10
US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOOMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEROF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; US-09-219-849-5

```

	Query Match	8.8%;	Score 452;	DB 4;	Length 960;
	Best Local Similarity	25.6%;	Pred. No. 1.be-21;		
	Matches 173; Conservative	60;	Mismatches 309;	Indels 134;	Gaps
QY	132 PTPKAEITTKGALTTPTKEPTPTTPEKEBASITPPEEPTTTIKSAPTTKEBPAP--TTTKS 183	:	:	:	:
DB	72 PCPAPCPSSRDGP--PCAPAGAG--PGSNDPPGAPG-PAPPCPSRDPGGAPGA 126	:	:	:	:
QY	190 APTTKEBAPTTTKEPA--PTTKEBAPTPTTKEBAPTTTKSAPTTKEBAPTTP--KKA 245	:	:	:	:
DB	127 GPPGSDGPCAPGAPGAPGSGSDGPGAGCGAPGSPGDCGAPGAPGPGSGSDGC 186	:	:	:	:
QY	246 PTTTKEBAPTTP--KEETPTTTKKPA--PTTKEBAPTPTKEPA----- 286	:	:	:	:
DB	187 P--PCAPGAGPGGSRDGPAGAPCAPPCPSRDPGGPGAGHAPGAPKGAMGAPGKGAHG 244	:	:	:	:
QY	285 -----PTMK-KKAPTTTP--KEBAPTPTKEPA--PTTKEBSPTTKEBAPTPTTKS 330	:	:	:	:
DB	245 PAGPKGAMGAPGKAPGAPGAGPGSGRDGPAGAPGAPPCPSRDGP--PCAPGP----A 299	:	:	:	:
QY	331 APPTTKEBAPTTTKSAPTTKPESPSTTTKEBAPTPTTKEBAPTTP--KKAPTTTKEBPAP 387	:	:	:	:
DB	299 GPPGSRDGP-----PCAPGAPGPGSGSRDGP--PCAPGAGCPGGSNDPPGP--PCAPGP 344	:	:	:	:

```

QY      388  TTP---KEPAPTTTUKKA-----PAPKEPAPTTTKEPAPTTPKUPTTEKLAPT 437
Db      349  AGPPGSRDPPGAPGAPGPPGSGDPPGAPGAPGPPGSRDGP-----PGAGPAPGPG 404
QY      438  PEKAPPT-TEPELAPTTPEEPPTTPEEPAPTTPKAA-ENTPEKAPPTP--KEPAPT 492
Db      405  SRDGPPEAPGAPGPPGSRDGP--PGAPGAPGSGDPPGAPGAPGPPGSRDGP 462
QY      493  TPKEPA-----PTTKEP-APTTPKGT-----APTTLKEPAPTTPKP-AREKLA 555
Db      463  GAHGAPGAGKGAHGAPGAGHAPGAPGAGHAPGAPGAPGAPGSGDPPGAPGAPG 522
QY      536  PTTTKEPTSTSDKAPAPTTPKGA-PTTKEPAPTTP--KEPAPTTPKGA-PTTLKBP 590
Db      533  PPGSGDPPGPA--GPPGPPGSGDPPGAPGAPGAPGSGDPPGAPGAPGSGRD 580
QY      591  APPTKKKAP-----KELAPTTTKGPTST--TSDKAPAPTTPKETAATTTKEPAPTTPK 642
Db      581  GP--PGAGPAPGPPGSRDPPGAPGAPGPPGSGDPPGAPGAPGAPGSGRDGP--PGA 636
QY      643  PAPPTP---ETPPPTTSEVSTPTTKEPTTIKSPDESPTPELSAETTPALEN-----S 693
Db      637  PGPAPGPPGSRDPPGAPGAPGAPGSGRDGP--PGAPGAPGSGDPPGAPGAPGAPG 695
QY      694  PKEBPVPTTKTPAATK 709
Db      696  SRDGPPEAGHGPAGPK 711

```

RESULT 11  
US-08-479-537A-5  
Sequence 5, Application US/08479537A  
Patent No. 5861381  
GENERAL INFORMATION:  
APPLICANT: CHAMON, Pierre  
APPLICANT: KIENY, Marie-Paule  
APPLICANT: LATHE, Richard  
APPLICANT: HAREVENI, Mara  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,537A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 90/13101  
FILING DATE: 23-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017733-025

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note="The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1 21
OTHER INFORMATION: /note="Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-5

```

```

Query Match      8.2% Score 423.5; DB 2; Length 1867;
Best Local Similarity 24.4%; Pred. No. 2.4e-19;
Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;

75 TSLVANKETTVETKETTNNKOTSTGKEKETSAAKETOSIEKTSAKD-LAPTSKVLAKPT 133
16 TLVTV-----VTGSGHASSTPGGEKETSATGRSSVSSTENKNAVSMTSVLSHS 65
134 P-KATTTKGP--ALTTPKEP-----TPPTPKREPASTPKPEPTTTIKSAP 176
66 PEGSGSTTGGDVTLPATPEPASGSAATWGODVTSVPTTRPALGSTPPPAHDVT---SAP 122
177 TTPKEPAPTTT-----KSAPTTPKPEP-----APTTPKEPAPTTKEPAPT 217
123 --DNKPAFGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 180
218 TKE--PAPTTT-----KSAPTTPKPEPAPTTPK-----KPAPTTPKEPAPTTPKEPTT 264
181 APDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 240
265 PKE--PAPTTKEP-----APTTPKEPAPTAPK-----KPAPTTPKEPAPTTPKEPAPT 311
241 APDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 300
312 TKEPSPPTPKPAPT--TKSAPTTTKEPAPT-----TKSAPTTPKPEPSPPTTK-----E 360
301 APDXRP--XPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 339
361 PAPTPPKREPAPTTPK-----KPAPTTPKEPAPTTPK-----EPAPTTTKKPAPTAPKEPAPT 412

```

```

Db 360 SAPDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 419
Oy 413 TPKETAP-----TPPKLAPTTPEKLAFTTPPEKAPPTTPPELAPTTPEPTPTTPPEAPPT 468
Db 420 SAPDXRPXGSTAPXAHGVT-----APDXRPXGSTAPXAHGVTSAPODXRPX-----PGST 471
Oy 469 TPKA-----AAPTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGT--APTTLKEPAP 523
Db 472 APXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRP--XPGSTAPXAHGVTSAPODXRPXG 530
Oy 524 TTP-----KKPAPELAPTTTKEPTSTTSKRAP--TTPKGTAPTTTPKEPAPTTP 571
Db 531 TAPXAHGVTSAPODXRPXGSTAP--XAHGVTSAPODXRPXGSTAPXAHGVT-----APDXR 585
Oy 572 KEAPPTPKG-----TAPTTLKEPAPTTP-----KKPAPELAPTTTKEPTSTSD 617
Db 586 PXPSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAP--XAHGVTSAPODX 644
Oy 618 KPAP--TTPK-----ETAPTTKEPAPTTP-----KKPAPTTPPETTPPTTSEVSTP 661
Db 645 RXPSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXG--TAPXAHGVTSA 702
Oy 662 TTPKEPT-----IHKSPESTPELSAEPTRKALENSKEPEGVPTTKTPAA-----707
Db 703 DXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSA 762
Oy 708 -TKPEMTTAAK--TTERDLRTTPETTTAAAPKMTKETATTEKTESKITATTGVT 763
Db 763 DXRPXGSTAPXAHGVTSAPODXRPXGST--APXAHGVTSAPODXRPXGSTAPXAHGVT 820
Oy 764 TTQDTTPFKITTLTKTLTTLAKVTTTKTTTITELMKNPEETAKPKDAINSKATTPK 823
Db 821 APDXRPXG-----STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTSAPODXRPX 868
Oy 824 KPTAPKPKPTTKKPKTPVRYKPKETTPPKMTSTPELNP--SRTAEMLQTTTPNQ 882
Db 869 -GSTAPXAHGVTSAPODXRP--XPGSTAPXAHGVTSA--PODXRPXGSTAPXAHGVTSAPODX 924
Oy 883 TP 884
Db 925 RP 926

```

```

RESULT 12
US-09-083-116-5
; Sequence 5, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREJVENT, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; City: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,537

```

```

: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/EP91/00835
: FILING DATE: 23-OCT-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/039,320
: FILING DATE: 04-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/403,576
: FILING DATE: 14-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Teskin, Robin L.
: REGISTRATION NUMBER: 35,030
: REFERENCE/DOCKET NUMBER: 017753-025
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1867 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FEATURE:
: NAME/KEY: Peptide
: LOCATION: 128..1727
: OTHER INFORMATION: /note="The amino acids spanning
: OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
: OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
: OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
: OTHER INFORMATION: repeats varies from 1 to 40."
: FEATURE:
: NAME/KEY: Peptide
: LOCATION: 134
: OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa
: OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
: OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
: FEATURE:
: NAME/KEY: Peptide
: LOCATION: 144
: OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa
: OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
: OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
: FEATURE:
: NAME/KEY: Peptide
: LOCATION: 147
: OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa
: OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
: OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
: FEATURE:
: NAME/KEY: Peptide
: LOCATION: 1..21
: OTHER INFORMATION: /note="Amino acids 1 to 21 are a
: OTHER INFORMATION: 21 amino acid precursor sequence."
: US-09-083-116-5

```

Query Match 8.2%; Score 423.5; DB 4; Length 1867;  
 Best Local Similarity 24.4%; Pred. No. 2.4e-19;  
 Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;

```

QY 75 TSLVANKETVETKETTNTKSTDKKETTSAKETOSIEKTSKD-LAPTSKVLAKPT 133
DB 16 TVLVI-----VTSSGHAASSPGGEKETSATQSRSSVSSTKKNAVSMSSVLSHS 65
QY 134 P-KAETTKGP--ALTTPKP-----TPPTTKPEASTTPKEPTTTIKAP 176
DB 66 PGGSGSTTGQDVTPLAPTEPAGSAAATWGDVTSVPTTRPALGSTTPPAHDTV--SAP 122
QY 177 TTPKRPAPTTT-----KSAPTTTPKP-----APTTTKPEAPTTTPKRPAPT 217
DB 123 --DNKRPAGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVT 180

```

```

QY 218 TKE--PAPTTT-----KSAPTTTPKEPAPTPK-----KPAPTTTPKEPAPTTKEPTPTT 264
DB 181 APDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 240
QY 265 PKE--PAPTTKEP-----APTTTPKEPAPTPK-----KPAPTTTPKEPAPTTTKEPAPT 311
DB 241 APDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 300
QY 312 TKEPPTTPKEPAPT--TKSAPTTTKEPAPT-----TSAPTTTPKEPPTTPK-----E 360
DB 301 APDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 359
QY 361 PAPTTTPKEPAPTTTPK-----KPAPTTTPKEPAPTTTPK-----EPAPTTTPKPAPTAKEPAPT 412
DB 360 SAPDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 419
QY 413 TPKEPAP-----TPPKKLTPTTPKEKLAPTTPKEPAPTTPEELAPTTPEEPPTTPEEPAPT 468
DB 420 SAPDXRPXGSTAPXAHGVTSA--APDXRPXGSTAPXAHGVTSA PDXRPX--PGST 471
QY 469 TPKA-----AAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGT-APTTLKEPAP 523
DB 472 APXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXG 530
QY 524 TTP-----KKPAPKEAPTTTKEPTTSTSDKRAP--TTPKGTAPTTTPKEPAPTTTP 571
DB 531 TAPXAHGVTSA PDXRPXGSTAP--XAHGVTSA PDXRPXGSTAPXAHGVTSA--APDXR 585
QY 572 KEPAPTTPKG-----TAPTTLKEPAPTTTP-----KKPAPKEAPTTTTPKGTSTSD 617
DB 586 PXPGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAP--XAHGVTSA PDX 644
QY 618 KPAP--TTPK-----ETAPTTTPKEPAPTTTP-----KKPAPTTTPPTPTTSEVSTP 661
DB 645 RXPGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAP--TAPXAHGVTSA 702
QY 662 TTPKEPTT-----IHKPDESTEPELSAEPTPKLLENSPKRPGVPTTKTPA----- 707
DB 703 DKRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA 762
QY 708 --TPKEPTTAKOK-----TTRDRLTTPETTTAPAKMTKEATTTETTESKITATTTQVTS 763
DB 763 DKRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA 820
QY 764 TTTQDTPPKITTKITTLAPKVTYTTTKKITTETEIMNKPDEETAKPKDRATNSKATTPK 823
DB 821 APDXRPXG-----STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTSA PDXRPX 868
QY 824 KPTKARKKPTSTKPKPTMPVRKPKTTTPRKMTSTMPLELNT--SRIAPAMLOTTTRPQ 882
DB 869 -GSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA--PDXRPXGSTAPXAHGVTSA PDX 924
QY 883 TP 884
DB 925 RP 926

```

RESULT 13  
 US-08-479-537A-2  
 Sequence 2, Application US/08479537A  
 Patient No. 5861381  
 GENERAL INFORMATION:  
 APPLICANT: CHAMBOU, Pierre  
 APPLICANT: KIENY, Marie-Paule  
 APPLICANT: LATHE, Richard  
 APPLICANT: HAREUVENI, Maira  
 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
 TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
 STREET: P.O. Box 1404  
 CITY: Alexandria

```

STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 - Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y - Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-2

```

Query Match 8.2%; Score 423.5; DB 2; Length 2035;  
Best Local Similarity 24.4%; Pred. No. 2.6e-19;

```

Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;
QY 75 TSLTVNKETTVERKETTNTKQSTDOKEKXTSAKETQSTIKTSAND-LAPTSLKVLAKPT 133
   |||
   |||
Db 16 TVLTV-----VYSGHASTPGCEKETATQSSVSPSTEKNAVMTSSVLSHS 65
QY 134 P-KAETTTKGP-ALTPKRP-----TPTTKEPASTTPKRPPTTKISAP 176
   |||
   |||
Db 66 PGSGSSTTGQDVTLPALPATEPASGSAATWGQDVTVPVTPPALGSTTPPAHDVT---SAP 122
QY 177 TTPKEPAPTTT-----KSAPTTPKRP-----APTTPKEPAPTTTPKEPAPT 217
   |||
   |||
Db 123 --DNKRAPGSTAXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVT 180
QY 218 TKE--PAPTTT-----KSAPTTPKRPAPTTPK-----KPAPTTPKRPAPTTPKRP 264
   |||
   |||
Db 181 APDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVT 240
QY 265 PKE--PAPTTKEP-----APTTPKEPAPTTAK-----KPAPTTPKRPAPTTTPKEPAPT 311
   |||
   |||
Db 241 APDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVT 300
QY 312 TKEPSPPTTPKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTTKEPSPPTTTK-----E 360
   |||
   |||
Db 301 APDXRP--XPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVT 359
QY 361 PAPTTTKEPAPTTPK-----KPAPTTTKEPAPTTPK-----EPAPTTTTPKRAPAPKEPAPT 412
   |||
   |||
Db 360 SAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVT 419
QY 413 TPKEPAP-----TPPKLTPPTPEKLAPTTPKEPAPTTPEELAPTTPEEPPTTPEEPAPT 468
   |||
   |||
Db 420 SAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVT 471
QY 469 TPKA-----AAPNTPKPAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 523
   |||
   |||
Db 472 APXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGS 530
QY 524 TTP-----KKRPKELAPTTTKEPSTISOKRAP--TTPKSTAPPTTPKEPAPTT 571
   |||
   |||
Db 531 TAPXAHGVTSAAPDXRPGSTAP--XAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXR 585
QY 572 KEPAPTTPKG---TAPTTTKEPAPTTT-----KKRAPKELAPTTTKEPSTISD 617
   |||
   |||
Db 586 PAXGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAP--XAHGVTSAAPDX 644
QY 618 KRAP--TTPK-----ETAPTTTPKEPAPTTT-----KKRAPTTTPPTTPPTSEVSTP 661
   |||
   |||
Db 645 RPXPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGS--TAPXAHGVTSAAP 702
QY 662 TTPKEPPTT-----THKSPDESTPELSAEPPTKALENSKEEGVPTTKTPAA----- 707
   |||
   |||
Db 703 DXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAP 762
QY 708 -TKPEMTTAKOK---TTERDLRTTPETTTAARPMKETAATTEKTTESKITTATTOVTS 763
   |||
   |||
Db 763 DXRPGSTAPXAHGVTSAAPDXRPGST--APXAHGVTSAAPDXRPGSTAPXAHGVT 820
QY 764 TTTQDTPPKITTKITTKITTLAPKVTTTKITTTTEINNKPEETAKKPRDRTATNSKATTPK 823
   |||
   |||
Db 821 APDXRPG---STAAPXAHGVT-----SAPDXRPGSTAPXAHGVTSAAPDXRPG--- 868
QY 824 KPRKAPKKTSTTKRPTMPRVKRPKTTTPRKMTSTMPLELNP--SLIAEMLQDTPRPNG 882
   |||
   |||
Db 869 -GSTAPXAHGVTSAAPDXRP--XPGSTAPXAHGVTSA--PDXRPGSTAPXAHGVTSAAPDX 924
QY 883 TP 884
   ||
   ||
Db 925 RP 926

```

RESULT 14  
US-09-083-116-2

Sequence 2, Application US/09083116  
Patent No. 6203795  
GENERAL INFORMATION:  
APPLICANT: CHAMON, Pierre  
APPLICANT: KIENY, Marie-Paule  
APPLICANT: LATHE, Richard  
APPLICANT: HAREJVENT, Mera  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,116  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/479,537  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1899  
OTHER INFORMATION: /note= "The amino acids spanning  
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat  
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
OTHER INFORMATION: repeats varies from 1 to 40."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,  
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA  
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
FEATURE:  
NAME/KEY: Peptide

LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
OTHER INFORMATION: 21 amino acid precursor sequence."  
US-09-083-116-2  
Query Match 8.2%; Score 423.5; DB 4; Length 2035;  
Best Local Similarity 24.4%; Pred. No. 2,6e-19;  
Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;  
QY 75 TSLTVNKTETVETKETTNNKQSTDDKEKTSKAKETOSIEKTSKAD-LAPTSKVLAKPT 133  
DB 16 TVLTV-----VTSCHASTPGCEKETSATQHSVSSSTEKMAVMTSSVLSHS 65  
QY 134 P-KAETTKGP--ALTTPKEP-----PTTPKEPASTTPKEPTTIKSP 176  
DB 66 PGSGSSTGGQDVTLAPATPEPASGSAATWGQDVTSVPTPALGSTTPAHVDT--SAP 122  
QY 177 TTPKEPAPTT-----KSAPTPKEP-----APTTPKEPAPTTKEPAPTT 217  
DB 123 --DNKPARPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 180  
QY 218 TKE--PAPTT-----KSAPTPKEPAPTTKEP-----KPAPTTKEPAPTTKEPTPT 264  
DB 181 APDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 240  
QY 265 PKE--PAPTTKEP-----APTTPKEPAPTPAK-----KPAPTTKEPAPTTKEPAPTT 311  
DB 241 APDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 300  
QY 312 TKEPSPPTTPKEPAPTT--TKSAPTTTPKEPAPTT-----TKSAPTTTPKEPSPPTTK--E 360  
DB 301 APDXRP--XPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 359  
QY 361 PAPTPKEPAPTTKEP-----KPAPTTKEPAPTTKEP-----EPAPTTTKKPAPTPAKEPAPT 412  
DB 360 SAPDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 419  
QY 413 TPKEPTAP-----TPPKKLPTTPKEKAPTPPEELAPTPPEPTTPPEEAPT 468  
DB 420 SAPDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 471  
QY 469 TPKA-----AAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT 523  
DB 472 APXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXG 530  
QY 524 TTP-----KKPAKELAPTTTPKEPTSTSDKAP--TPPKGAPTPTPKEPAPTT 571  
DB 531 TAPXAHGVTSAPODXRPXGSTAP--XAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRP 585  
QY 572 KEPAPTTPKPG--TAATTLKEPAPTT-----KKPAKELAPTTTPKEPTSTSD 617  
DB 586 PXPSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODX 644  
QY 618 KPAP--TTPK--EAPTPTPKEPAPTT-----KKPAPTTPPTTPPTTPPTTPPTTPPT 661  
DB 645 RPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODX 702  
QY 662 TTPKEPTT-----IKSPDESTPELSAETTPALNSPREPEVPTTKPAA----- 707  
DB 703 DXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSA 762  
QY 708 -TKPEMTTADK--TTERDLTPPTTPPTTPAPKMTKEATTEKTESKITATTTATTTGTS 763  
DB 763 DXRPXGSTAPXAHGVTSAPODXRPXGST--APXAHGVTSAPODXRPXGSTAPXAHGVT 820  
QY 764 TTODTTPFKITTLKTTTLAPKVTTKKTTITTEIMNKPEETAPKODAINSKATTPKPO 823



Db 821 APDPRPG-----STAPXAGVT-----SAPDXRPXPGSTAPXAGVTSAPDXRXP- 868  
QY 824 KPTAPKRPSTKPKTTPRVKPKTTPPKMTSTMPELNPT-SRIAEMLOTTPPNQ 882  
Db 869 -GSTAPXAGVTSAPDXRXP--XPGSTAPXAGVTSAPDXRXPXAGVTSAPDX 924  
QY 883 TP 884  
Db 925 RP 926

## RESULT 15

US-08-276-967-2  
; Sequence 2, Application US/08276967  
; Patent No. 5851817  
; GENERAL INFORMATION:  
; APPLICANT: Hardy, Daniel M.  
; APPLICANT: Garbers, David L.  
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of  
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,967  
; FILING DATE: Submitted Herewith  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: US/SD:418\KIT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-787-1400  
; TELEFAX: 713-789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2476 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-276-967-2

Query Match 8.1%; Score 419.5; DB 2; Length 2476;  
Best Local Similarity 32.5%; Pred. NO. 5.7e-19;  
Matches 163; Conservative 43; Mismatches 177; Indels 119; Gaps 31;

QY 406 PKPAPPTPKETAPTPPKLPTTPPKLAPTPPEKAPATTPPEELAPTPPEEPPTTPPEP 465  
Db 313 PSETSVSEKPVAPTE---KPTVSEIYITPTEKPMVHMEKPIVHT--EKPTVPT-EKP 365  
QY 466 APPTKAAAPNPKPBA---PTPKPAPPTPKPAPATTPKETAATTPPKGAPATTLKP 521  
Db 366 TPTPKSTVPT--KPTVFKETLPPE-GPTVPAP-RPTTPREGPAPPKG--PTVLTE- 418  
QY 522 APPTPKKAPKELAPTTTKPEPTSTSDKAPATTPKGAATP--TPKPAATTPKPAATPK 580  
Db 419 -----WPTSHTEKSTVHTKELILPTGKSTIPTKPMVPTKRT-----TTP- 458  
QY 581 GATPTLKEPAPATTPKAPKELAPT--TKGPTSTT---SDKAPATTPKETAATTPKPEPA 636

Db 459 -TERTIIPAKPTV---PIEKPMVPTERTTIPERTTIPTEKPTVPTKLTVPPT--EKPI 512  
QY 637 PTPPKKAPATTPPETPPP-----TTSEVSTPT-----TKPEPTTIHKSPDESTPELSAEP 686  
Db 513 VPTEKPIVPTKHTIPTKELIVLTERTTIPTERTTIPTEKPIVPTPEKPSVPT-EKPTVPT 571  
QY 687 PKALENSPKPEGVPTTKTAATKPEMTTAKDKTTERDLRTTPEPTTAAPKMTKETATTT 746  
Db 572 -----EERTTPTKELTVPPT--ERTTTPKRTTPTTPTTPTTPTTPTTPTTPTTPTTPT 613  
QY 747 EKTTESKITATTTQVSTTTODTTPPKITTLTKTTLAPVTTTKKTTTTEIMNPEEIA 806  
Db 614 ERTT-----TPTRTTPTERTT---IPTKKT-----VPEKTIPT-----ERTI 652  
QY 807 KPKDRATNSKATTPKQKPTKAPKKTSTKPKRTMGRVKKPRTTTPPKMTSTMPELNPT 866  
Db 653 AP-----TTPQP-SPTLVPTQPAVVMST-----SATTVPTPTTIASCP---PN 693  
QY 867 SRIAEMLOTTPRNPOTPKSKL 888  
Db 694 AHFERCACPVSQ-SPTPNCCL 714

Search completed: April 26, 2002, 16:17:05  
Job time: 216 sec

Mon Apr 29 08:35:34 2002

us-09-556-246-1\_copy\_200\_1167.ra

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:22 ; Search time 82.85 seconds  
(without alignments)  
931.380 Million cell updates/sec

Title: us-09-556-246-1\_copy\_200\_1212  
Perfect score: 5416  
Sequence: 1 VKDNKKNRKKRPKPPVY.....SPIDVPTRCNCEGKFFFK 1013

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062.5	19.6	3020	2	A43932 mucin 2 precursor, S-layer protein - cyst germination s
2	942	17.4	1664	2	T18262 hypothetical prote
3	853	15.7	1489	2	T31108 glucan 1,4-alpha-g
4	802	14.8	1274	2	T16251 extensin-like prot
5	789.5	14.6	1367	1	S48478 nascent polypeptid
6	784	14.5	1188	2	S49915 mucin 1 precursor,
7	776	14.3	2187	2	T30826 high molecular mas
8	772	12.4	1344	1	A35175 hypothetical prote
9	671	12.4	1151	2	T18535 elastin class 1 p
10	659.5	12.2	1229	2	T25697 hypothetical prote
11	633	11.7	489	2	T11622 hypothetical prote
12	633.5	11.7	7962	2	T18346 gene LF3 protein -
13	632	11.7	761	2	C84672 larval glue protei
14	629	11.6	3507	2	T34513 hydroxyproline-ric
15	626.5	11.6	990	2	T51618 nucleolar phosphop
16	625	11.5	6642	2	T29757 protein UNC-89 - C
17	623.5	11.5	839	2	T75518 mucin MUC5B, trach
18	622	11.5	3570	2	T15025 hypothetical prote
19	620.5	11.5	971	2	T19431 hypothetical prote
20	607.5	11.2	801	2	T29018 hypothetical prote
21	607.5	11.2	924	2	S27923 gene LF3 protein -
22	605	11.2	379	2	S50125 larval glue protei
23	559.5	10.3	350	2	S22456 hydroxyproline-ric
24	556.5	10.3	856	2	T16543 hydroxyproline-ric
25	556.5	10.3	875	2	S23760 polyphenolic adhes
26	555.5	10.3	1630	2	A53577 asclites sialoglyco
27	553	10.2	2232	2	T34434 hydroxyproline-ric
28	551	10.2	620	2	S06733 hydroxyproline-ric
29	545	10.1	873	2	A47283 calphostin - fruit

30	542	10.0	369	2	S20500 hydroxyproline-ric
31	538.5	9.9	416	2	TJ0465 extensin precursor
32	533	9.8	1087	1	OPMSH neurofilament trip
33	532	9.8	756	2	T27642 hypothetical prote
34	530.5	9.8	1162	2	A70557 exo-alpha-sialidas
35	530	9.8	865	2	A47282 calcium-binding pr
36	522	9.6	328	2	JC00985 hydroxyproline-ric
37	522	9.6	1459	2	T32271 hypothetical prote
38	518	9.6	813	2	S70795 vsta protein precu
39	518	9.6	866	2	T45462 membrane glycoprot
40	518	9.6	1072	1	A37221 neurofilament trip
41	512.5	9.5	867	2	T45463 membrane glycoprot
42	509.5	9.4	662	2	A45155 mucin FIM-C.1 - Af
43	508	9.4	1832	2	T31113 mucin-like glycopr
44	505	9.3	700	2	A54641 interspersed repea
45	504.5	9.3	606	2	A43427 neurofilament trip

#### ALIGNMENTS

RESULT 1

A43932  
mucin 2 precursor, intestinal - human (fragments)  
N:Alternate names: mucin SMUC-41  
C:Species: Homo sapiens (man)  
C>Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 05-Nov-1999  
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A>Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t  
A:Reference number: A49963; MUID:94132002  
A:Accession: A49963  
A:Molecule type: mRNA  
A:Residues: 1-639 <GU1>  
A:Cross-references: GB:L21998  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A>Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
A:Reference number: A45106; MUID:93016075  
A:Accession: A45106  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 626-1895 <GU2>  
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396  
A>Note: sequence extracted from NCBI backbone (NCBI:116706)  
A:Accession: B45106  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 2037-3020 <GU3>  
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398  
A>Note: sequence extracted from NCBI backbone (NCBI:116698)  
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,  
J. Clin. Invest. 88, 1005-1013, 1991  
A>Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym  
A:Reference number: A43932; MUID:91358717  
A:Accession: A43932  
A:Molecule type: DNA  
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864  
A>Note: sequence inconsistent with the nucleotide translation  
A>Note: sequence extracted from NCBI backbone (NCBI:55749, NCBI:55750)  
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A>Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl  
A:Reference number: A33532; MUID:89197956  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GU4>  
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874  
A:Experimental source: Intestine  
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991  
A:Title: Human bronchus and intestine express the same mucin gene.  
A:Reference number: A61257; MUID:91086481  
A:Accession: A61257  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>  
A:Experimental source: bronchus  
R:Xu, G.; Huan, L.; Khattai, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner  
Biochem. Biophys. Res. Commun. 183, 821-828, 1992  
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t  
A:Reference number: P00328; MUID:92198477  
A:Accession: P00328  
A:Molecule type: mRNA  
A:Residues: 2328-2468 <XUG>  
A:Cross-references: GB:M6523  
A:Experimental source: small intestine  
A:Accession: P00329  
A:Molecule type: protein  
A:Residues: 2328-2342,'K',2344-2354 <XUG1>  
C:Genetics:  
A:Gene: GDB:MUC2  
A:Cross-references: GDB:120203; OMTM:158370  
A:Map position: 11p15.5-11p15.5  
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von  
C:Keywords: glycoprotein; intestine; tandem repeat  
A:Zn:2766-2834/Domain: von Willebrand factor type C repeat homology <WMC>

[illegible]

QY	513	-----	512
Db	1820	NISCRATMYDPVIGOLGOTVCDSVGLICKNEODKREGVIPAFAFLCINTEINQOCECV	1879
QY	513	TAPPTLKEPAPTPPKKPAKELAPTT--TKRPTSTTSDK-----APT--TPKGAAPTPKE	565
Db	1880	TQPTITM---TTTTENTPTPIITTTTTVTPRPTPTSTOSPNGLAPRPTPISTTTVTVPTR	1936
QY	566	PAPTPKEAPPTPKGIAPTTLKEAPPTPKKPAKELAPTTTKRPTST--TSOKPAPTPR	624
Db	1937	PTPTGQTPR--TPPTTTTTTTTTVTPRPTGQTPPTVLITTTTTTTTTPTPTSTIKSTVTP	1995
QY	625	KETAPTPKBPAPTPKRP-----APPTPEPTPTSEVSTPTTKR-----EPPTIHK	672
Db	1996	ITTTTTVTAAPTPPTGQTPPTMIDISTTTTTVTPRPTPTGTSIGRPHNTSTAIAELITSNP	2055
QY	673	SPDESTPELSAEPTPKALENSPEKPGVPTTKTPAATKREPMTTAAKDTENDLRTPEPT	732
Db	2056	PPESSTPQTSKSNSSPPLTESTLLSTPLPALEMTSTAPSPSTPAAPTTSGGHTLSPSPST	2115
QY	733	TAAPKMTKREATTITTEKTEKTIATTTQVTSITTDPTFKITT---LKTITLAPKVYTT	789
Db	2116	TTPSPGTPRTGTTT--GSSAPPSPTVOTTTTSAMVPTPTPLSTPSIIRTTGLRYPBSV	2173
QY	790	KKITITTEIMNKPE-----TAKPKDRATNSKATPR	820
Db	2174	LICCVLNDTYAAEGEEVYNGTYGDTCYFVNCISLCTLEFYNMNSCSTSPPTPSPK--STP	2232
QY	821	KPKPRTAPKPKPTSTKKPKTMPVRKKPKTTPTRPKMTSTMPLENPSTRIAEAMIQTTTPR	880
Db	2233	TPSKRSPSPKPTGTPRCPDPCDFDPR-----QENETMMLCDOCFM--ATCKY	2278
QY	881	NOTPNSKLVEVNP-----KSDAGC-----ABEGTPTHMLL-----	910
Db	2279	NNNTPEIKVCEBERPMPETCSNGIQPVAYVEDBDCCHMHECDOYCTGMGD--PHYTFDGLY	2337
QY	911	-----RPHVEMPEVTPDMQYTLPRVNOGIIIN	937
Db	2338	YSTQGNCTYUVEIETISPSVD-----NFGVYID	2364

```

RESULT      2
T18262
S-layer protein - Clostridium thermocellum
C.Species: Clostridium thermocellum
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C.Accession: T18262
R:Fujino, T.; Begun, P.; Albert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A.Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A.Reference number: Z18447; MUID:93209931
A.Accession: T18262
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1664 <FUJ>
A.Cross-references: EMBL:X67506; NID:q296879; PID:q296881; PIDN:CAA7841.1

Query Match          17.4%; Score 942; DB 2; Length 1664;
Best local similarity 33.0%; Pred. No. 1..3=33;
Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;

QY    128 VLAKPTP-KAEETTKGPAITTPKEPTTPKESASTTPKEPTPTTIKSAPTTPKAPATT 186
      V I Q A P I K A A D E D I P D T P S D E P T P -----D E P T P S ---D E P T S D E P T P S D
Db     758 WICQAPITKAADDEIPDTDPSEDEPTP-----DEPTPS---DEPTSDSEPTPIPTDP 804

QY    187 TKSAPTPKEPAPTTPKCAPTTKEPA---PTTKEPAPTTPKSAPTTEKAPATPPK 242
      : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     805 EPTPSSTPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEETPEEPIPTDP 864

QY    243 KPAPTPKCAPATTPKEPTTPTKKCAPATTPKEPATP-TPKCAPAPAKPKAPATTPKEAP 301
      I I : : : | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

```

Db 865 SDEPTSPDEPTSS--DEPTPS--DEPTP--SDEPTSPSEPTPEIPTDTPSDEPTSPDEPTP 919  
QY 302 TTPKEPA--PTTKESP--TPKEPAPTTKS--APTTPKEPAPTTKSAPTTKESPPT 357  
Db 920 SDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 11  
QY 358 TKEPAPTTKKEPAPTTKKPAPTTKKEPAPT--TPKEPAPTTKKPAPTTKKEPAPTTKPE 416  
Db 969 SDEPTSS--DEPTPS--DEPTPSDEPTSPSEPTPEIPTDTPSDEPTSPDEPTSPD 1019  
QY 417 TAPTTKKTLPPTPEKLAPTTKPEKAPPTPEELAPPTPEEPPT--TPKEPAPTTKKAAAP 475  
Db 1020 -----EPPTSPD--EPPTSPD-----EPPTSPDEPTSPSEPTPEIPTDTPSDEPT 1064  
QY 476 TPKKEPAPTTKKEPA--PTTKKEPAPT--TPKEPAPTTKGATPLTKKEPAPTTKKPAK 532  
Db 1065 TSPDEPTSPDEPTSPDEPTSPDEPTSPSEPTPEIPTDTPSDEPTSPDEPTPS--DEPTPS 1122  
QY 533 ELAPTTKKEPT--STSDKAPATTPKGTAPTTKEPAPTTKKEPAPTTKGATPLTKKEPA 591  
Db 1123 D-EPPTSPDEPTSPSEPTPEIPTDTPSDEPTSPDEPTPS--DEPTPS--DEPTSPDEPT 1175  
QY 592 PT--TPKKAPEKELAPTTGPTSTSDKAPAT--TPKTAATTTKEPAPT--TPKKAAPT 646  
Db 1176 PSEPTPEP-----IPTDTPSDEPTSPDEPTSPDEPTSPD--EPPTSPDEPTSPSEPTPEIPT 1229  
QY 647 TPEPTPTSPSEPTPTTKPTTKPTTHKSPDEPTPELSAEPPTPALENSPKPEGVPTTKPA 706  
Db 1230 TPSPDEPTSPD--EPPTSPDEPT-----PSDEPTP--SDEPTP--SDEPTPEP--IPTDTPSD 1277  
QY 707 AKRPEKTTAKKTTERDLITPEPTTAAKMKETATTEKTESKITATTOYSTTT 766  
Db 1278 EPPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPSEPTPEIPTDTPSDEPTPS 1337  
QY 767 ODTPEKITTLLTTTLARVTTTKTITTTIMNKPEEAKAKDANATKATTPKQ--KP 825  
Db 1338 DEPTSPDEPT-----PSDEPTSPDEPTSPDEPTSPDEPTSPDEPTP 1372  
QY 826 TPAKRPSTTKKPTMPRVKRPKPTTPPKMTS-----TMPPLNPTSRLA 870  
Db 1373 SEPTPEPTPTT-----TPPTSTPTSSGSGSGSGSGGGGGGTVPSPTPTPS--- 1424  
QY 871 EMLQTTTPRNPOTNSKLVENPKSEDAAGAGETPHMLLRPHVMEVTPMDYLP 927  
Db 1425 -----KPTSTPAPTIE--EPPTSPDVGAIGGHRAYLRG-----PDGSPRP 1465

RESULT 3  
T31108  
cyst germination specific acidic repeat protein precursor - Phytophthora infestans  
C:Species: Phytophthora infestans (potato late blight agent)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T31108  
R:Goerhardt, B.  
submitted to the EMBL Data Library, April 1998  
A:Reference number: Z20986  
A:Accession: T31108  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1489 <GOE>  
A:Cross-references: EMBL:AF061185; NID:93851513; PID:93851514; PIDN:AACT2308.1  
C:Genetics:  
A:Gene: car90

Query Match 15.7%: Score 853; DB 2: Length 1489;  
Best Local Similarity 32.3%: Pred. No. 7.6e-30;  
Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

QY 9 TTKKTPRPVVDAGSGIDNGFKVTTPDTSTT-----OHNKVSTSPKITTTAKPIMR 62  
Db 338 TPVATPEKRYDVEETTYTTESTYATPTKSEINAPTERMHYAHIEKPCDTEVTAKAPTEET 397

QY 63 PSLP-----PNSDT-----SKETSLVKNKETTIVETKETT---TNKQSTND 100  
Db 398 TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 457  
QY 101 GKEKTTSAKETSIOIKTSKADLAPSKVLAKPRAEETTKCPALTTKREPTTPKKEA 160  
Db 458 TEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 514  
QY 161 STTPKEPTPTTKSAPTTPKKEPAPT--TKSAPTTPKKEPAPT--TPKKEPAPTTPKKEPAPT 217  
Db 515 EETTPKEPTPEET--TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 573  
QY 218 TKEPAPTTKKSAPT-----TPKKEPAPTTPKKAAPTTPKKEPAPTTPKKEPTTPKE 267  
Db 574 TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 632  
QY 268 P-----AP-----TPKKEPAPTTPKKEP-----APAPKPAAPTTPKKEPAPTTP 304  
Db 633 PTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 692  
QY 305 KEPAPT--TPKPSPTTPKEP-----APT--TKSAPT-----TPKKEP-----A 339  
Db 693 TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 752  
QY 340 PT--TKKSAPT--TPKPSPTTPKEP-----APTTPKKEPAPTTPKKAAPTTPKKEPAPT-- 388  
Db 753 PTEATTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 812  
QY 389 TPKEPAPTTKKPAATKAPKAPATTPKETAAPTTPKPLIPTPEKILAPPTPEKAPATTPPEE 448  
Db 813 TPYEPTTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 870  
QY 449 LAPT--TEETPTPTPEE-----APT-----TPKKAADNTPKKEPAPTTPKKEP- 489  
Db 871 YAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 930  
QY 490 -----APTTPKPAAPTTPKETAAPTTPKGTAPT--LTKKPAAPTTPKKAAPT--ELAPT--TT 539  
Db 931 EETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 990  
QY 540 KEPTSTSDKAPATTPKGTAPTTKKEPAPTTPKKEP-----APTTPKGTAPT----- 585  
Db 991 YAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 1050  
QY 586 -----LTKKPAAPTTPKKEP-----APKE--LAPT--TPKGPSTSDKAPAPTTP 624  
Db 1051 EETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 1110  
QY 625 KETAPTTKKEPAPTTPKKAAPTTPPETPTSPSTSEVSTPTTKKEPT--THKSPDEPTPELS 682  
Db 1111 YAPAEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 1169  
QY 683 AEPTPKA--LENSPKPEP-----VPTTKPAATKPEMTTAAKDTTERDLRPTTP-----ETT 732  
Db 1170 TEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 1229  
QY 733 TPAKMKETATTEKTESKITATTOYSTTQDIT-----PEKITT--LKT 779  
Db 1230 TVAP-----TEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 1284  
QY 780 TTLAPKVTTKKITTTPTEIMNKPEETAKPKDRAT--NSKATTPRP--OKPTAKRPKPT--- 833  
Db 1285 TTVAPTEATTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 1344  
QY 834 -----STKKP--KTPMRVAKRPKPTTPPKMTSTMBELNPTSRLAEMLQTTTPR--N 881  
Db 1345 PAESTSTVSTKPCONTLEETDEPTDEPTD--PSDEPTDEPTDEPTDLPDSEPTSCDN 1402  
QY 882 QTPNSKLVENPKSEDAAG 899  
Db 1403 OGINGIGVENKRVYNNAG 1420

RESULT 4  
 T16251  
 hypothetical protein F35A5.1 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
 C:Accession: T16251  
 R:Leinbach, D.  
 submitted to the EMBL Data Library, January 1996  
 A:Description: The sequence of *C. elegans* cosmid F35A5.  
 A:Reference number: 218485  
 A:Accession: T16251  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1274 <LEI>  
 A:Cross-references: EMBL:046675; NID:g1166613; PID:g1166621; PIDN:AA52641.1; GSPDB:GN00  
 A:Experimental source: strain Bristol N2; clone F35A5  
 C:Genetics:  
 A:Gene: CESP:F35A5.1  
 A:Map position: X  
 A:introns: 1272/2

Query Match 14.8%; Score 802; DB 2; Length 1274;  
 Best Local Similarity 28.0%; Pred. No. 1e-27;  
 Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

QY 6 KNRKK-KP-----KPVVDEAGSLDNGD-----FKVTPPTSTQHNKVST 50  
 DB 274 KNTKKKMPWEDTEVEEVEKPEPVKAPVLKKDPAPAAKADSPSKAAPKVEPS 333  
 QY 51 PKTTAKPI-NPRSLPNSDTSKETSINVKETVETETTTNKOTSTDKETTSK 109  
 DB 334 SPVPPPTPKNPVKKKPKPWEVDEDAEVEKKPSAEKPTVLKREPESSSTPSDDS 333  
 QY 110 ETOSIEKTSADLAPTSTKVLAKPTPKAE-----TTTKGA-----LTPP 148  
 DB 394 PKAAAVKPRDSSPKKAPFLQADPKAEVPPVKNPKVKYKPEVDEDEDEVEEVEKQ 453  
 QY 149 KEPTPTP-----KEPASTTKEP-----TP-TTKSAPTPKPEAPTTTSAPTPEAP 199  
 DB 454 EAPAKKTPVLKREKPEAKDTAKPATSKTPTPKKDPVKRDPSSPKVAKKPSAQA 512  
 QY 200 TTKKEPA-----PTPKPEAPTTTKPEAPT-----TTKSAPTTP 233  
 DB 513 TPVKNVYKKMRPPEDETPADVSKPTDAKTPSLAKKDPAPAKSLKKAUTKAPAK 572  
 QY 234 KEP-----APTTKKAPPTPK-----PAPTTKKEPTPTTKPEAPTTKPEAPTTK-EP 283  
 DB 573 RDPSPKKVAPTAPBEKTPVLAKKEPAGPDSKTEPEKSPKRDSPKAVPAKVPKTEV 632  
 QY 284 APAPKKAP-----TTKEPAPTTPKPEAPTTTKPE-----SPTTPKEPAPTTKSAPT 333  
 DB 633 APAVKKPEISPKDTAPKAPNSVVP-PIPVKNPKVKMKPWWDDAPAKPSLPE 691  
 QY 334 TTKPEAPTTKSAPTTP-KEPSEPTTKPEAPTTPK-EPAPTTPKKAP--TTKPEAPT 388  
 DB 692 PEKK-TPVLAKKAPTKDESAADPVSGPSSKDKLAKKAPVPRDPSPKAVPIKAPK 750  
 QY 389 TKPEAPTTKKAPPA-PKEPAPTTPKPEAPTTPKKLTPTTP-----EKLAPTTPKPA 442  
 DB 751 T--EVPAVVKKPEPAKSDPSKKAK-AEENSP--VPEPTPKVKNPVKMKPWWDDA 805  
 QY 443 PTPPEELAPTTPEP-----TPTPPEPAPTTPKAAAPNTPKPEAPTTKPEAPTT-PKE 496  
 DB 806 PAEAVNVPPEEKTPVLAKKTPVKNRDPSPKAVPAKPSRTAPVSVKKEPEVSKPE 865  
 QY 497 PAPTTPKEAPTTPKGAPTTLKEPAPTTPKKAPKELAPTT-TPKEPTSTTSKAPATTP 555  
 DB 866 PSPKAPNSVVP-----PIPVKNPVKMK-KPWEDEDETEVEKKPSE--PEKKTPLA 918  
 QY 556 KGATPTPKPEAPTTPKPEAPTTPKGAPTTLKEPAPT-----TPKKAP-----TEL 603  
 DB 919 K-KEPKPKD-APKVAAPKRDSPSKAVPE--KEPAKVAAPKRLSLKKAIPANTQEA 974

QY 604 APTTKGPTS-----TNSDKP-----APTTPKET-----APTTKEPAPTTKAPAP 646  
 DB 975 PTPVKNPVKMKKPPWEDDEDAEPAVSAPEPEKTPVLAKKAPAKPRDP-----SPKAPAV 1031  
 QY 647 TPETPTTSEVSPPTTKPEPTTIKSP-----DESTPELSA-EP--TPKALENSPKKEPG 698  
 DB 1032 AAK-PDPKRIPEP-PTPVKNPVKMKKPPWEDDEDESEPAVSAPEPEKTPVLAKKAPAKPA 1089  
 QY 699 V-----PTTKPAAT-----KPEMTTADKTKTERLRTPTPT--TAAPK 737  
 DB 1090 TKPDSEAAADPVSGPTSKDPLSKKAPVEKPKPTTDPKDDKIKESPAKPEKAPPAAPK 1149  
 QY 738 MKKETATTEKTESKITATTOVSTTTQDTTPPKITTLKTTTLAPKVTTKKTTTTE 797  
 DB 1150 KKKPWWDDPDEPDEDFYPAKSKKPTIEDPADPLG-----GPKTKDPK----- 1193  
 QY 798 INKKEETAPKPKDRATNSKATTPKPKOKPTKAPKPTSKPKTMRVRKPK----- 848  
 DB 1194 -LNKKAPAKKPTK-----PKKEVSKPEPKKPTPEPKP-AADKKKPPWEDDPEDEPE 1243  
 QY 849 ---TTPPKKMTSTMP 861  
 DB 1244 ADETPAPKPKPDTEDP 1259

RESULT 5  
 S48478  
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)  
 N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01  
 C:Species: *Saccharomyces cerevisiae*  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 12-Nov-1999  
 C:Accession: S48478; A26877; B26877; S27281; J06123  
 R:Rowley, K.  
 submitted to the EMBL Data Library, October 1994  
 A:Reference number: S48478  
 A:Accession: S48478  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <ROW>  
 A:Cross-references: GB:Z47047; EMBL:Z38061; NID:9603997; PID:9763364; GSPDB:GN00009;  
 R:Yamashita, I.; Nakamura, M.; Fukui, S.  
 J. Bacteriol. 169, 2142-2149, 1987  
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.  
 A:Reference number: A91831; MUID:87194600  
 A:Accession: A26877  
 A:Molecule type: DNA  
 A:Residues: 1-242 <YAM>  
 A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525  
 A:Accession: B26877  
 A:Molecule type: DNA  
 A:Residues: 762-1331 <YA2>  
 A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526  
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Clares, M.G.; Jimenez, A.  
 FEBS Lett. 239, 179-184, 1988  
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Sacc*  
 A:Reference number: S27281; MUID:89031230  
 A:Accession: S27281  
 A:Molecule type: DNA  
 A:Residues: 1-31 <PAR>  
 A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552  
 R:Lambrecht, M.G.; Bauer, F.F.; Marmur, J.; Preorius, I.S.  
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
 A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseud  
 A:Reference number: J06123; MUID:96323237  
 A:Accession: J06123  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <LAM>  
 A:Cross-references: GB:U30626; NID:g1304386; PIDN:AA49609.1; PID:g1304387  
 C:Genetics:  
 A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
 A:Cross-references: MIPS:YIR019C; SGD:S0001458  
 A:Map position: 9R

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
F:5-21/Domain: transmembrane #status predicted <TM1>  
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match	14.6%;	Score 789.5;	DB 1;	Length 1367;
Best Local Similarity	28.4%;	Pred. No. 3.8e-27;		
Matches 301;	Conservative 104;	Mismatches 483;	Indels 171;	Gaps 47

Qy	35	TPBDJSTTQHNKVNVSPIKTTAKPLNP-----RSLPBNDSPTKESLTVNKEETVE	86
Db	236	STSESSTSTSSSESTSSSTSTTAAPTPTTSCTEKKEPPTPTTSCTEKKEPPLPHHDTPC	295
Qy	87	TKETTTTNKQSTDQKEKTTSAKEQOSLEKTSANDLAPTSVLAKPRPKAETI--TKGPA	144
Db	296	TKKKTTSK-TCT---KKTTPVPPTPS--STTESSSAV-----PPSSSTTESSSAV	344
Qy	145	LTTPKE-----PPTTKEBASTTKEPPTTIKSAF---TPKBPAPTITKSAPTT	194
Db	345	TSSSTESSAPVPPTSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPPTSSSTTE	404
Qy	195	KEBPATTTKEBPATTTKEBPATTTKEBPATTTKSAPTTKEBPATTPKBPATTPKEBP	254
Db	405	SSSAPPTS---TTESSSAVTS---STTESSAPVT---SSTTESSAPVTSTTES	453
Qy	255	TTTKEPTP---TPKBPATTTKEBPATTTKEBPATTAKKAPATTPKEBP---TTPKB	307
Db	454	SSAPPTPTSSSTTESSAPVT---STTESSAP-VPTPTSSSTTESSAPVTSSTTESS	509
Qy	308	APTITKEPPTTKEBPATTTKSAPTTTKEBP---TTTSAPTTKEPSPPTTKEBP	363
Db	510	AVPTPTPS--STTESSAPAPITPSSSTTESSAPVTSSTTESSAPVPPTSSSTTESS	568
Qy	364	----TPKBPAPT-TPKBPATTTKEBPATTTKEBPATTTKKBPAPKBPATTPKETA	418
Db	569	VTSSTTESSAPVPPTSSSTTESSAPVT---PSSSTTESSAPAPPTSSSTTESSA	624
Qy	419	PTTPKLPPTTPEKLPATTPKBPATTPPELAPT-TPEEPTPTTPEBPAP---TTPA	472
Db	625	PVT-----STTESSAP-VPTPTSSSTTESSAPPTPSSSTTESSAPVPPTSSSTTESS	679
Qy	473	AAPTPKBPATTTKEBPATTTKEBPATTTKETAPT-TPKGAPITLKEBPATTPKBPAP	531
Db	680	SAPVT-----SSTTESSAPVT---SSTTESSAPVPPTSSSTTESSAPVPPTSSSTTE	731
Qy	532	KELAPTITKEPTSTTSOKRAPATTPKGAPITTPKEBP--TPKBPAPT-TPKGAPITL	587
Db	732	SSSAPVPPTPS--STTESSAPAVTSSTTESSAPVPPTSSSTTESSAPVPPTSSSTTESS	790
Qy	588	KEBPATTTKBPAPKELAPTTTKGPTSTTSOKRAPATTPKETAPTTPKEBP-----TPPK	642
Db	791	SAPVPPTSSSTTESSAPVPPTSSSSNIISAPSTPSSSTTESSSVAPPTSSSTTESS	850
Qy	643	PAPTPTPEPTTSEVSTPT-----TPKBEPTTIHKSPDESTEPELASAPPKALENSKPE	697
Db	851	SAPVSSSTTESVAPVPPTSSSSNIISAPSSI--PFSSTTESSTG-IYTPSSSKTP	906
Qy	698	GVPPTTKBPAPAKPMTTAAQDKTENDLKTPTPTTAAAPKMTKEVATTTTEKTESKITAT	757
Db	907	GSQETSVSSTETTIVPTTKTTSVTPPTTTTIVCSTGTSAGETTSGCSPKVATTT	966
Qy	758	--TTQVSTTTPQOOTTPPKITTLKTTLTLPVNT---TKTITTTT--EIMKKPEETAKPDR	811
Db	967	VPTTTTTSVTTSTTTTITTTVCSTGTSAGETTSGCSPTTTTTPVCSTSPETA-----	1021
Qy	812	ATNSKATTPKPKOKP-----TKAPKPTSKTKPKMPVRKPKTTPPR	854
Db	1022	---SESPTTSTPTVTVTVTVTEYVSTRKRGELITTTTVTKNIPITTYLTITAAPP-	1077
Qy	855	KMTSTMBELNPTSKIAIAMAQITTRPNQPNBSKLVEVNPSSDAGAGEGPHMLRPHV	914
Db	1078	-SVTVNFPPT-ITTVTVCSTG-----NAGETTSGCSPKVATTTVPPTST	1122

QY 915 FMEPTPEMDLPRVPNGIILNMLSDENICNGRPVDCITLLTRNGTLVAFRGHEFMM 974  
Db 1123 GCGEITTTATLTVTAATTTVTTESSTGNSA-GTTTTGTTKSVPTT-----YVTL 1175  
QY 975 SPSPSPPARITEVCGIPSEIDTVTRC---NCEGKT 1009  
Db 1176 AFSALVTEATN-----AVPTTTT--TECSAATNAGET 1207

RESULT 6  
S49915  
extensin-like protein - maize  
C:Species: Zea mays (maize)  
C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
C:Accession: S49915  
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.  
submitted to the EMBL Data Library, June 1994  
A:Description: Pex genes: pollen-specific genes with extensin-like domains.  
A:Reference number: S49915  
A:Accession: S49915  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1188 <NUB>  
A:Cross-references: EMBL:Z24465; NID:g600117; PIDN:CA84230.1; PID:g600118

Query Match	14.5%	Score 784;	DB 2;	Length 1188;
Best Local Similarity	29.9%;	Pred. No. 5.7e-27;		
Matches 230;	Conservative	59;	Mismatches 359;	Indels 122;
			Gaps	33;
QY 13	PTPKPPVNDVDEAGSGLDNGDFKVTYPTDSTSTQHNKSTSPKRTITAKPINRPSL----	PPN 68		
Db 460	PTPHSPPAD-----DYVPTPTPVPGKSPATSPSPQVQPPAASPTPPLVLSLSPQ	510		
QY 69	SOTSKEISLTVNKKETTVETKEITTNNKOTSTIDGKRTTTSAKTOSIEKTSADKLAPTSKV	128		
Db 511	APVG-----SPPPVKTTSPAPDIG----SPSPPPVSVV	541		
QY 129	-----LAKPTKRAETTTTGPALTTPKEEPTPTTKBPASTTPKEEPTTKASPTTKKEPA	183		
Db 542	SPPPVKSPPPAPVAGSPPEPKSPPPAPVAVASPPPVKSP--PPPLVASPPPVKSPSP	599		
QY 184	PTTTSASPTTP--KEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT	240		
Db 600	PPAPVASPPPVKSPPTPTPVASPPPPAVASPPPMKSPPTPTPVSSPPPEKSPPPPP	659		
QY 241	PKKPAPTTKEPAPT-----PK-----PTPTTKEPATTTKEPATTTTKEPATTPAKKP	291		
Db 660	PAKSTPPPEEPTPTPTSVKSSPPPEKSLPPTLISPPPOKEPTPTPTSPKRP--PSSPKRP	718		
QY 292	APTTPKEPATTTTKEPATTTTKEPSPTTTPKEPATTTTSAPTTKEPATTTTKEPATTTT	351		
Db 719	SP--PKREVSSPPQIP-----KSSPPPAVSSPPPTPVSSPALAVSSSPSKSSP----	768		
QY 352	EPSPTTKEPATTTTKEPATTTTTPKKPATTTTKEPATTTTKEPATTTTTPKKPATTPAKKEBP	411		
Db 769	PPAPLSSPPAPQVKSSPPVQVSSP--PPAKSSPLPAVSSPPQVEKTSPPAPLSSPP	827		
QY 412	TTPKETAP---TTP---KKLTPTTPEKLAPTTPEKAP-----TTPTELAPTTPEEPT	458		
Db 828	LAKSSPPHHVVSSPPPVKSSPPPAFVSSPPPLPKAPASPAHVVSSPEVVKPSTP--PA	885		
QY 459	PTTPEEPATTPKAAPPTPKP--PATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT	517		
Db 886	PTTYISP--PSEPKSSPPTPVSLPPLPVKSSPPPAVSSP--PMTKSSPPVVVSSPPT	943		
QY 518	LKE---PA-----PTPKK---PAKTELAPTTTKE---PSTTSDKRAPTTPKGTAPPTP	563		
Db 944	VKSSPPAPVSSPATRKSSPPPAVNLPPPEVSSSPPTPVSSPPA--PKSSSPAP	1000		
QY 564	KEPATTTTKEPATTTTKEGAPTTLKEPATTTTTPKKPAKELAPTTTGTSTTSDKAPT	622		

Db 1001 MSSPPPEVKSPPAPVSSPPPVKSPPPPAPVSSPP--PPVKSPPPPAPVSSPPPV 1057

QY 623 -TPKEAPATTTPKEAPATTTPKPPAPATTPEPTTSEVSTPTTTTKEPTTTHKSPDESTPEL 681

Db 1058 KSPPPAPVSSPPPVKSPPPPAPVSS--SPPPVKSPPPPAPVSSPPPVKSPPPPAPVSS 1115

QY 682 SAEPTKALENSPKEGVPPTTKTPAATKPEMTTATKDKTTERDLRTTPT 731

Db 1116 SPPAP-----VKPPSLP-PPAPVSSPPPVVPPAPPKKEOSLPPAPAS 1158

RESULT 7

T30826

nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse

N:Alternate names: alpha-NAC protein

C:Species: Mus musculus (house mouse)

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 24-Nov-1999

C:Accession: T30826

R:Yotov, W.V.; St-Arnaud, R.

Genes Dev. 10, 1763-1772, 1996

A:Title: Differential splicing in of a proline-rich exon converts alphaNAC into a muscle

A:Reference number: 220889; MUID:96312450

A:Accession: T30826

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2187 <YOT>

A:Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AA18732.1

C:Genetics:

A:Gene: Naca

A:Map position: 10

A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3

A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ

C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 14.3%; Score 776; DB 2; Length 2187;

Best Local Similarity 27.7%; Pred. No. 2, 2e-26;

Matches 288; Conservative 122; Mismatches 389; Indels 240; Gaps 49;

QY 34 VTTPDSTTOHNVSTSPKITTAKPINDRPSLPPNSDTSE-----TSLTVNKETT 84

Db 806 VOSKVPDPSMDVPTSPKTSATAPV-PK-----DTSATLSKSVPAVTSLSPPKAPV 857

QY 85 VETETTTNNQSTSDGKEKTSKENSIEKTSAKDLAPTSKVLAPTPAEETTGP 144

Db 858 APSNEATVPEIPTSILKNALAAATPKETLATSIPKVTSP-----PQKTKSVSLKAP 913

QY 145 LTT-----PK-----PTPTTPKEPASTT-----PREPTPTT 171

Db 914 MTSKATPIASKDVSPSQPFKEVPLQHVPTSPKSPVSDTLGALTSPPKGP-PAT 972

QY 172 IKSAPTPKE-PAPTTKSAPTPK-----EPAPTTKEAPATTTPKEPATT 221

Db 973 LAETPTTKKSPKPAASAKRTPATSPSEGVAVPLEIPCSKKAATAAPKSSATSSSR 1032

QY 222 APTT--TTSAPT-----TPKEAPATTTPK-PAPTTPKKEPAPTP 257

Db 1033 APKTAVSKVPSKGVAVPLEISLPKTSKSNATPGEKSSAPKRSPTAPKE---TPP 1089

QY 258 KEPTPTTKEPAPTTKEPAPTTKEPAPTTAPKAPATT--KREPAP---TTPKEPAPTT 311

Db 1090 GGVTAVPPEISLPKETQONNTPNLSLAASSOKSPKSVKETPPGCVTAMPLEIPSA 1149

QY 312 TKESPPTPK-----EPAPTTKSAPTTTKEPAPTTT-KSAPTPK-----EPSTTT 357

Db 1150 QKAPTAAPKQIPREDVAVTLLAGSPLSPKASKTAAPKAPVPSGVAVSGEISPS 1209

QY 358 TKEPAPTTKEPAPTTKPPAPTT--PRE-----PAPTTKEPAP---T 396

Db 1210 KTSSTAAKENSATLPPRSKTAAPKETPATSEGVTAVPSELSPSPPTPASKGVPT 1269

QY 397 TTKKAPTAAPKAPATTTPKETAATTTPKLTPTTPKLAATTPEK-PAPTTPEELAPTTPE 455

Db 1270 LTPKGAANLAE-SPASPKVKPKTAPAEETSTPT-----SPQIKPVAGCPKEASATPPS 1322

QY 456 EPTPTT-----PEEPAPTTPKAAAPNTPKEAPATTPE-PAPTTTKEPAPTTTPKETA 506

Db 1323 KTKPTAVPKETSAPSEGVTAVPLEIPSPKAPKTAAPKETAAPAS--PEGATTAPOIP 1380

QY 507 PTPPGTAPTTTKEPAPTT-----KKPAPKELAPTTTKEPTS-- 544

Db 1381 PPKRKSAGKAGSKE-TPPTPSSEGVTAAPLEIPSSKTSKMAKSKETLVPSSKLSQT 1439

QY 545 ----TTSKAPATTTPKGAATTPTPKEAPATTPE-PAPTPPGTAPTTTKEPAPTTPKPA 599

Db 1440 VGPKESTLEGATAVALEIPSSHKKAPKTVPKQVLTSPK-DAPTTLAE-SPSSPK-A 1496

QY 600 PKEAPATTTKGPTSTTSKAPATTTPKEAPATTTPKEPAPTTTPKPPATPEPP-----PT 654

Db 1497 PTAAPPSER-VTTPPEKPA-TPQKASGTASKVPVPAETQEVAVSSRETPVTPAVPV 1554

QY 655 TSEVSTPTTK-----EPTTHKSPDESTPELSAEPKALENSPKE-PGVPTTKTPAA 707

Db 1555 KNPSSHKTSKTELKEAPATLPPSPKSPKIPSSKKAAPT--SAPKEFPAPSISK-PVT 1611

QY 708 TPEMTTAKDKTTERDLRTPEPTTAAPKMTKETATTTEKTESKITATTQVSTTQ 767

Db 1612 T-SLAQTAPPSLOKAPSTTIKEMLAAPV-----LPVSSKSPAPPARASASLSPATAAP 1665

QY 768 DTTPPKITLTKTTTLAPKVTTKTTTTEIMNKPETAKPPDRATNSKATPPQKPPK 827

Db 1666 QTAPEKATTTIPSCKKAATETPEITASTLSGAKREISE-----TSVSLVLMSSP----- 1716

QY 828 APAKPTSTKPKTKTPRVKRPKTPPRKMTSTMPELNPTSRJAEMALQTTTPNPNK 887

Db 1717 -PKKASSKRASTLP-----ATLPLSKEASVLS-----PTATISSGK 1752

QY 888 LVEVNPKSEDAAGAGEPTP 906

Db 1753 DSHISPV-S-DACSTGTTTP 1770

RESULT 8

A35175

mucin 1 precursor, repetitive splice form A [validated] - human

N:Alternate names: breast carcinoma-associated Dp3 antigen; core protein Kp39; epista

nceatic mucin; polymorphic epithelial mucin (PEM)

N:Contents: mucin 1 precursor; epithelial tumor antigen splice form; mucin 1 precursor

C:Species: Homo sapiens (man)

C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 02-Jun-2000

A:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218;

R:Lightenberg, M.J.L.; Vos, H.L.; Gemmissen, A.M.C.; Hilkens, J.

J. Biol. Chem. 265, 5573-5578, 1990

A:Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene

A:Reference number: A35175; MUID:90202794

A:Accession: A35175

A:Molecule type: mRNA

A:Residues: 1-952,1033-1344 <LIG>

A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35806.1; PID:g182124;

A:Experimental source: splice form A

A>Note: GenBank entries HUMEPIS1A and HUMEPIS1A2 present only the amino- and carboxyl

A:Accession: B35175

A:Molecule type: mRNA

A:Residues: 1-19,29-952,1033-1344 <LIG2>

A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129;

A:Experimental source: splice form B

A>Note: GenBank entries HUMEPIS1B and HUMEPIS1B2 present only the amino- and carboxyl

R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burt

J. Biol. Chem. 265, 15286-15293, 1990

A:Title: Molecular cloning and expression of human tumor-associated polymorphic eplth

A:Reference number: A35886; MUID:90368715

A:Accession: A35886

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19,29-992,1033-1344 <GEN>

A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870







```

QY 257 PKEP-TPPTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE- 314
Db 325 THRPVGTSTN--FDLATEPFAKSEDKMTLSKATATETTTQOTETV-DGPEKETIKV 301
QY 315 -----PSTTTKEPAPTTKSA-----TTTEPAPTT-----TTKAPTTPKESPPTTK 359
Db 382 SIELPITVLETTSTSTASRESDFHTLLKVTADSDSTESATVVKPNEETTTK 441
QY 360 E-----PAPTT-----EPAPTTPKKAP-----TTTK 383
Db 442 SHVPRKTKKGYKVPKLELSEDEPTEIT-KAPHKELLEKTYHFLVSDNFARSEAK 500
QY 384 E-----PAPTTKEPAP-----PAPTTKEPAP-----PKPAPTT-----PKETAP 419
Db 501 ENDYNDLADNYHREAKEPTTTEESSTTEBVTTEBPANTGNPPTENPTTQPTSTAE 560
QY 420 TTPKKLLPPTPEKLA---PPTPEKAPTTPEELAPTTPEPTTTPPEAPTTPKAAPN 476
Db 561 STTALPPTTEOTVTEPTEPATAKSTATO---KPTTQSVST---EKSTTKKA---S 610
QY 477 TPKEPAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAP 536
Db 611 TTEB---PTTDEPTTTT---ESSTGKATPELSTSEETTTTELKITT-----GS 657
QY 537 TTTKEPTSTSDKAP---TPPKGAPTTPEKAPTTPEKAPTTPEKAPTTPEKAP 592
Db 658 TTTTEPPTTAIFAEASTGIIITDEETTSSTPTEITSKE--IVGESAIOTQSVSVS 715
QY 593 TTP-----TPKAPKELAPTTTKGP-----TTT 613
Db 716 STPRLPERKKAIVNKFKNHLEVLKKEKRLKKESTSGSDSSTTTVAVENIDEVTT 775
QY 614 TTSDBAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAP 670
Db 776 TEKEVQOTVPTTEKSTTQOEBETTTTTEKTSKTTTEKPTTSATTEITTSPTSP- 834
QY 671 HKPDSSTPELSAPPTPKALENSPKKPGVPTTKPATPEMTTAKDKTTERDLTPE 730
Db 835 -----TEST-----IVDTSSATTEESSTAETTTSAE---TSE 865
QY 731 TTTA-----APMTKETATTTTEKTE 751
Db 866 TTTSEAPITGSPENTALOSSSKSENESSAEKGRDRDFVPRKHAFTVVKAPATTSA 925
QY 752 SKTATTTTQ-VNST---TQDPTPEKITTLKTTTLAPKYT---TTKITTITTEINMKPEE 804
Db 926 VAASSTTPTPTTTEKSTTLETTPIEATTLNTEVGAFTVGAPEVDETITNTELSK--- 982
QY 805 TAKPKRATNSKATTPKPO-----KPKAPKKTPTSTKPKPTMPRVKRP 847
Db 983 -----INNTQISOPKPTDISKTALSLISGLIGSFTKAPMAPTI----- 1022
QY 848 KTTTPRKMTSTMPELN-----PTSRIAE 872
Db 1023 HTTTDAFVTATBALSNGSDKKIIDEAQTDEIRRA 1059

```

## RESULT 11

11622  
 extensin class 1 precursor - cowpea  
 C:Species: Vigna unguiculata (cowpea)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 17-Nov-2000  
 C:Accession: T11622: S54155  
 R:Arsemljevic-Maksimovic, I.; Broughton, W.J.; Krause, A.  
 M01. Plant Microbe Interact. 10, 95-101, 1997  
 A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.  
 A:Reference number: 217301; MUID:97155574  
 A:Accession: T11622  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-489 <ARS>  
 A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937

A: Experimental source: sub species Red calona  
 R: Arsemljevic-Maksimovic, I.; Broughton, W.J.; Krause, A.  
 submitted to the EMBL Data Library, April 1995  
 A: Description: A class of root-hair specific extensins involved in rhizobium/legume 1  
 A: Reference number: S54155  
 A: Accession: S54155  
 A: Status: preliminary  
 A: Molecule type: mRNA  
 A: Residues: 326-489 <AR2>  
 A: Cross-references: EMBL:X86030; NID:g791149; PID:g791150  
 C: Geneticks:  
 A: Gene: Ext26G  
 C: Superfamily: hydroxyproline-rich glycoprotein  
 C: Keywords: glycoprotein; hydroxyproline  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-489/Product: extensin class 1 #status predicted <MNT>

Query Match 11.7% Score 633; DB 2; Length 489;

Best local similarity 31.2% Pred. No. 7.5e-21;

Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 7;

```

QY 179 PKEPAPTTKSAPTTKEPAPTTTKEPAPTTKEPAPTTTKEPAPTTTKEPAP 238
Db 39 PKOTPPYYNAPPYYKSP-----PSPSP-----PPYVHKPPYYKSP 83
QY 239 TTPKKAFTTPEKAPTTKEPTTTPKAPATTKEPATT-----KEPAPTKKAPPT 295
Db 84 PSPSPPPYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 141
QY 296 PKEPAPTTPEKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 355
Db 142 YKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 187
QY 356 TTTKEPAPTTPEKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 415
Db 188 YYYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 247
QY 416 ETAPTPPKLPTTPEKLAFTTPEKAPTTPEELAPTTPEEPTTTPPEAPTTKAA 475
Db 248 PPPYYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 307
QY 476 NTPKEPAPTTPEKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 535
Db 308 PSPSPPPYYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 366
QY 536 PTTTKEPTSTSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 595
Db 367 -----KSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 422
QY 596 KKPAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 653
Db 423 SPP-----PPYYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 475

```

## RESULT 12

138346  
 elastic titin - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
 C:Accession: I38346  
 R:Labelf, S.; Kolmerer, B.  
 Science 270, 293-296, 1995  
 A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.  
 A:Reference number: A57430; MUID:96026330  
 A:Accession: I38346  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-7962 <RES>  
 A: Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427  
 C: Geneticks:  
 A: Gene: GDB:TTN  
 A: Cross-references: GDB:127867; OMIM:188840



QY 448 ELAPTTPEERTPTPEEPART-----TPKAAP-NPKEPARTPKBPARTPKBPARTTP 502  
DB 421 KLPVKKPRPPIYSPVKKPRVHKPRPIISPKRPPVHKPRPIIKRPPVKKPRPPR 480  
QY 503 KETAPTPKGTAPTLKEBAPT--TPKKAPKELATTTKEPTSTSDKRAPTPKGTAP 560  
DB 481 TTSPPVOP-----PVOKRPPPTSPVKKRPIOKRPT---PTYSPIKRPVKKRPTPY 532  
QY 561 TTPKEBAPTTPKEBAPT-TPKGTAPTLKEBAPT--TPKKAPKELATTTKGTSTSD 617  
DB 533 SPPIK-PPVHKPRPTYSPIKRPPIIKRPIIYSPRIKRPVHKPRPTYSPIKRP 591  
QY 618 KAPPTPKETAPTPKEBAPTTPKBPAP-TPETPTTSEVSTPTTK-EPPIHKSP 674  
DB 592 VHKPRPTYSPIK-PPVHKPRPTYSPIKRPVHKPRPTYSPIKRPVHKPR 647  
QY 675 DSTPELSAEPKALENSPKRGPVPTTKAPATKREMTTAKDKTTEDLTTEPTTA 734  
DB 648 ---TPYSPPIKRPVOKRPPPTYSBPVKKRPVQLRP-TPYSPVKKRPPVOPPTPTYS 703  
QY 735 APKMTKETATTEKTESKITATTTQVSTTQDTPP 771  
DB 704 PVKKRPPVOPPTPTYSPIKRPVOPPTPTPTSP 740

RESULT 14  
T34513  
hypothetical protein ZK783.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34513  
R:Avellio, A.; Vaudin, M.  
submitted to the EMBL Data Library, August 1994  
A:Description: The sequence of C. elegans cosmid ZK783.  
A:Reference number: 221536  
A:Accession: T34513  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3507 <FAV>  
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GNO0021; CESP:ZK783.1  
A:Experimental source: strain Bristol N2; clone ZK783  
A:Gene: CESP:ZK783.1  
A:Map position: 3  
A:Insertions: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;  
3504/1

Query Match 11.6%; Score 629; DB 2; Length 3507;  
Best local similarity 24.1%; Pred. NO. 7.2e-20;  
Matches 252; Conservative 158; Mismatches 416; Indels 218; Gaps 45;

QY 10 KKKPKPPVVDGAGSLDN-GDFKVT-----PDTSTQHKKVSTSPKTTA-----K 57  
DB 1942 KKQPKREKIEIDEMSSSSNGQEKPTTKGIYSSISATSSSESTAEPIVTTISISTTK 2001  
QY 58 PINPPSPPPN---SDTKETSLVANKETV-EKETTTNKQSTSDKEKTTSAKQOS 113  
DB 2002 DMTSSKS-PEVNTMSSESEVSTSSKSTASSETVSTPSSSESEAPLNSPATTEV 2060  
QY 114 IEKTSAKLAP-----TSKVLAKPPKAEKTTGKALTTPKKEPTPTTPKEPASTTKETP 169  
DB 2061 ITESVSKTTKESSSEITVKKLSKSPVESSVSKSPSPS-TPSOSVSTVPESTKS 2119  
QY 170 TTIKS-APTPKEBAPTTPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTTKS 228  
DB 2120 TVLSSEAEVTSNPTREVHT-SSETPKSLASSTGDTNSTPSSSLASVASTAPETS 2178  
QY 229 APPTPKBPAPTPPKBPAPTPPK-EPAPTPPKBPPTPTTPPKBPAPTPKEP-APTPKEBAPT 286  
DB 2179 ASVAVKLSLSPDVOSPTSTTFDATESSTVQASSETSVKSTSEPSHVTKLSTSS 2238  
QY 287 APKBPAPTPPKBPAPTPPKBPAPTTKEPSPPTP--KEBAPTTPKSAPTTTKEBAPTTPK 344

DB 2239 NPSVSPVTSKSTPTPE-----STEOPTSTTPGQSLTTPKNSSEVLTTSEPIVLS 2293  
QY 345 SVP-----TPKEPSPTTKBPAPTPKEBAPTTPKBPAPTPPKBPAPTPPKBPAPTP 397  
DB 2294 LSPDVOSSTTPNNLSSESTVE-----TPKTSSEVLSNBESEPTBPAPTLSPDLSTTTN 2349  
QY 398 -----TKKBPAPK-----EPAPTPKEBAPTTPK 427  
DB 2350 NLSOSTVSTEDRSISSESEKPT-SAPELVTSVTHVASSPDVPRES--SEPDLLTG 2406  
QY 428 TTPKEIAPTPPEKBPAPTPPEELAPTPPEEPPTPEEP-----APTPPAAPN 476  
DB 2407 SSTEINPEASSKQITSSPTPTPTTASSEPTKSTMSDLSTNVLSSESTTPPSS-- 2464  
QY 477 TPKEBAPT-----PKBPAP-----TPKEBAP-----TPKEBAPTTP 510  
DB 2465 --KSPVSSSTEGISVVTSTERSKVPBESTISSVLEEDLTKTPSPILEETVASETSEPLT 2522  
QY 511 KGTAPTLKEBAPTTPKBPAPKELAPTTTKEPTSTSDKA-----PT----- 553  
DB 2523 EDSLTVSRHIELTTSSEVKESESTTSSSESKPSEAPGILTSVVPPTSSVSLTA 2582  
QY 554 -----TP--KGTAP--TPK-----EPAPTPKEBAPTTPKGTAPTLKEBAPT 593  
DB 2583 SEIEATSTNTPKQGRPTTSPKSLVASTSPSTVTSSEPSSESTKRTTSTVSTPTPT 2642  
QY 594 TPKKAPKELAPTTKGT-STSDDKBPAPTPPKETAPTPPKBPAPTPPKBPAPTPPTPP 652  
DB 2643 EETTSSELILTAAPSKPTESTSESEAPTPPAKSEKPNVSTSKSTENVEVTSQ 2702  
QY 653 PTTSEVSTPTTKETPTTHKSPDESTP--ELSAEPTPALE--NSPKRGPVPTTKPAT 708  
DB 2703 SGLSESTMSST-----SEPETAAPVTVSSEASTLLENSTSSP-----TSSEASV 2752  
QY 709 K-----PEMTTAAKDTTERDLRTTPETTTAAPKMTKETATTEKTESKITATTT 759  
DB 2753 KLSLFPESITSEAVTVASR---APAEITMSSEHREISVSSPSPELPLSTVSPN 2808  
QY 760 -----QVSTTQDTPP-FKIT-----LKITLAPKVTYTKTTTTEIMKP 802  
DB 2809 VVTASSISEPILSSVSSSTPRVRLITGPDLLIVSVPSHGNNRQNTASSV--P 2865  
QY 803 EETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKMPBPVBPK--TPTPRKMTST 859  
DB 2866 SNSNSPILPSES-LTTPQRPPTTTTAKPATTSGKRKPSIOPPAEKFTTPAP----- 2918  
QY 860 MPELN-----PTSRIAEAMLOTTT 878  
DB 2919 PPSNGYGEETNOEEQVSTTT 2942

RESULT 15  
I51618  
nucleolar phosphoprotein - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 24-Sep-1999  
C:Accession: I51618; S57757  
R:Calins C.; McStay B.  
J. Cell Sci. 108, 3359-3347, 1995  
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp  
A:Reference number: I51618; M01D:96019267  
A:Accession: I51618  
A:Molecule type: mRNA  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-990 <CAI>  
A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921  
C:Gene: xNopp180  
C:Superfamily: nucleolar-cytoplasm shuttle phosphoprotein  
C:Keywords: phosphoprotein



Mon Apr 29 08:35:37 2002

us-09-556-246-1\_copy\_200\_1212.rpr

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:50 ; Search time 47.4 seconds

(without alignments)  
783.575 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1212  
Perfect score: 5416  
Sequence: 1 VKDNKKRRTKKRTPKPVV.....SPIDVFTRCNCEGKTEFFK 1013

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwisProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	1168	21.6	5179	1	MUC2_HUMAN
2	942	17.4	1664	1	SLPI_CLOM
3	789.5	14.6	1367	1	AMYH_YEAST
4	651	12.0	1255	1	MUC1_HUMAN
5	556.5	10.3	875	1	FPI_MYED
6	555.5	10.3	2700	1	ZAN_HUMAN
7	551	10.2	620	1	EXTN_TOBAC
8	533	9.8	1087	1	NFH_MOUSE
9	530.5	9.8	1162	1	TCNA_TRYCR
10	530	9.8	865	1	CPN_DROME
11	518.5	9.6	872	1	FPI_MYTCO
12	509.5	9.4	662	1	MUC1_XENLA
13	503.5	9.3	1970	1	RPB1_HUMAN
14	498.5	9.2	1970	1	RPB1_MOUSE
15	497	9.2	831	1	NFH_RAT
16	493.5	9.1	467	1	RPB1_CRIGR
17	488.5	9.0	826	1	SSP2_PLAYO
18	488	9.0	1020	1	NFH_HUMAN
19	475.5	8.8	267	1	EXTN_MAIZE
20	471.5	8.7	5376	1	ZAN_MOUSE
21	468.5	8.7	634	1	HMP1_CANAL
22	467	8.6	817	1	VRL1_YEAST
23	463	8.5	2142	1	BAT2_HUMAN
24	454	8.4	797	1	VGLX_HSEB
25	454	8.4	1161	1	V39P_YEAST
26	448.5	8.3	670	1	VG50_HSV1
27	442.5	8.2	751	1	FPI_MYGA
28	439.5	8.1	1083	1	T2D3_HUMAN
29	432.5	8.0	1439	1	XP2_XENLA
30	432	8.0	3164	1	TEGU_HSV1
31	426	7.9	1275	1	TRX2_HUMAN
32	424.5	7.8	1125	1	MAP4_MOUSE
33	424	7.8	307	1	SGS3_DROME

34	424	7.8	1251	1	YOU3_CAEL
35	422.5	7.8	1794	1	VAV1_SCHPO
36	422.5	7.8	3421	1	TEGU_HSVB
37	419.5	7.7	2476	1	ZAN_PIG
38	419.5	7.7	2774	1	MAPA_RAT
39	419	7.7	1185	1	DRPL_HUMAN
40	415	7.7	1229	1	N121_HUMAN
41	411.5	7.6	907	1	VEP3_EBV
42	410	7.6	1183	1	DRPL_RAT
43	407.5	7.5	3256	1	K167_HUMAN
44	405.5	7.5	2517	1	NCR2_HUMAN
45	405	7.5	1411	1	TCOF_HUMAN

## ALIGNMENTS

RESULT	ID	SEQUENCE	STANDARD	PRT	5179 AA.
1	MUC2_HUMAN	002817: 014878;			
AC	01-JUN-1994 (Rel. 29, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).				
GN	MUC2 OR SMUC.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine;				
RX	MEDLINE=94132002; PubMed=8300571;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;				
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.				
RT	Identification of the amino terminus and overall sequence similarity				
RT	to prepro-von Willebrand factor";				
RL	J. Biol. Chem. 269:2440-2446(1994).				
RN	(2)				
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.				
RC	TISSUE=Colon;				
RX	MEDLINE=93016075; PubMed=1400449;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.W., Lagace R.E.,				
RA	Kim Y.S.;				
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located				
RT	both upstream and downstream of its central repetitive region.";				
RL	J. Biol. Chem. 267:21375-21383(1992).				
RN	[3]				
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.				
RX	MEDLINE=91358717; PubMed=1885763;				
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,				
RA	Petersen G.M., Kim Y.S.;				
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays				
RT	and polymorphism.";				
RL	J. Clin. Invest. 88:1005-1013(1991).				
CC	-1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND				
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A				
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIONS				
CC	AGENTS AT MUCOSAL SURFACES.				
CC	-1- SUBUNIT: MULTIMERIC.				
CC	-1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,				
CC	BRONCHUS, CERVIX AND GALL BLADDER.				
CC	-1- PFM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR				
CC	INTRACHAIN DISULFIDE BONDS (BY SIMILARITY).				
CC	-1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND				
CC	VARIABLE AMONG DIFFERENT ALLELES.				
CC	-1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT				
CC	OF SILKMOIR HEMOCYTIN.				
CC	-1- SIMILARITY: CONTAINS 2 WFEC DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).				

CC	-	-	-
DR	EMBL; L2198;	AAB95295.1;	-
DR	EMBL; M74027;	AAA59875.1;	-
DR	EMBL; M94131;	AAA59163.1;	-
DR	EMBL; M94132;	AAA59164.1;	-
DR	MIM; 158370;	-	-
DR	InterPro; IPR000359;	Cys_knot.	
DR	InterPro; IPR000561;	EGF-like.	
DR	InterPro; IPR002400;	GEF_cysknot.	
DR	InterPro; IPR001007;	VWFC.	
DR	InterPro; IPR001846;	vwd.	
DR	Pfam; PF00007;	Cys_knot; 1.	
DR	Pfam; PF00094;	vwd; 4.	
DR	PRINTS; PR00438;	GFCYSKNOT.	
DR	SMART; SMO0214;	VWC_def; 2.	
DR	SMART; SMO0011;	VWC_def; 2.	
DR	SMART; SMO0216;	VWD; 4.	
DR	PROSITE; PS00022;	EGF_1; UNKNOWN_1.	
DR	PROSITE; PS01185;	CTCK_1; 1.	
DR	PROSITE; PT01225;	CTCK_2; 1.	
DR	PROSITE; PS01208;	VWFC; 2.	
KW	Glycoprotein;	Repeat; signal.	
FT	SIGNAL	1	POTENTIAL.
FT	CHAIN	21	MUCIN 2.
FT	DOMAIN	1401	APPROXIMATE REPEATS.
FT	REPEAT	1401	1.
FT	REPEAT	1417	2.
FT	REPEAT	1433	3.
FT	REPEAT	1433	4.
FT	REPEAT	1445	5.
FT	REPEAT	1449	6.
FT	REPEAT	1472	7A.
FT	REPEAT	1479	7B.
FT	REPEAT	1495	8A.
FT	REPEAT	1518	8B.
FT	REPEAT	1534	8C.
FT	REPEAT	1557	9A.
FT	REPEAT	1573	9B.
FT	REPEAT	1597	10A.
FT	REPEAT	1612	10B.
FT	REPEAT	1635	11A.
FT	REPEAT	1656	11B.
FT	REPEAT	1652	12.
FT	REPEAT	1676	13.
FT	REPEAT	1684	14.
FT	REPEAT	1700	15.
FT	REPEAT	1716	16.
FT	REPEAT	1732	17.
FT	DOMAIN	4815	VWFC 1.
FT	DOMAIN	4924	VWFC 2.
FT	DOMAIN	5075	CTCK.
FT	DISULFID	5075	BY SIMILARITY.
FT	DISULFID	5089	BY SIMILARITY.
FT	DISULFID	5098	BY SIMILARITY.
FT	DISULFID	5102	BY SIMILARITY.
FT	DISULFID	? 5159	BY SIMILARITY.
FT	CARBOHYD	163	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	423	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	670	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	770	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	894	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	1139	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	1154	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	1215	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	1230	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	1246	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	1787	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	1820	N-LINKED (GLCNAC . . .) (POTENTIAL).

[illegible]

QY	492	TPPKAPAPL-TPKKEAPLTPPKCAPLTLKKEAPLTPPKKAPKAPKELAPLTPPKKEPTSTMSXK	550
Db	1950	TTTTVTPRTPTGQTPTTTPTTTTTPVTPRPTPTGQTPTTTPTTTTTPVTPRPTPTGQT	2009
QY	551	AP-TPPKCAPLTPPKKEAPLTPPKKEAPLTPPKCAPLTLKKEAPLTPPKKAPKAPKELAPLTPPK	609
Db	2010	TTTTPTPTTTTTPVTPRPTPTGQTPTL-TPVTTTTTTPVTPRPTPTGQTPTTTPTTTTTPV	2068
QY	610	GPTSTTSOKRAP-TPPKKEAPLTPPKKEAPLTPPKKAPLTPPELTPPTTSVSTPTTTKEPT	668
Db	2069	TPPTPTPTGQTPTTTPTTTTTPVTPRPTPTGQTPTL-TPPTTTTTPVTPRPTPTGQTPT	2127
QY	669	THKSPDSTRELSAEPTRKALENSFKER-----GVPT-KTKPAKPEMTTAK	717
Db	2128	T---TPITTTTTPVTPRPTPTGQTPTTTPTTTTTPVTPRPTPTGQTPTTTPTTTTTPV	2184
QY	718	DKTERDLR---TTP-ETTPAPKMT--KETATTEKTEKSKITATTTQVSTTTQDT	769
Db	2185	PTPTPTGQTPTTTPTTTTTPVTPRPTPTGQTPTTTPTTTTTPVTPRPTL-TGQTPTPT	2243
QY	770	TPFKTTLTKTLTLAKVTT-KKITTTEIMKKPEETAKPKRATNSKATTPPKPOKPTKA	828
Db	2244	TP-TTT--TTTTPTPTPTGQTPTTTPTTTTTPVTPRPTPTGQTPTTTPTTTTTPV	2299
QY	829	P-KRTSTKPKPTMRVARKPKTTPPRKMTSMPELNP--SRAIALLOTTR-PNOT	883
Db	2300	PTPTPTGQTPTTTPTTTTTPVTPRPTPTGQTPTTTPTTTTTPVTPRPTPTGQTPTTT	2359
QY	884	PNSKLVEVNPKSEDAGAEGETPHMLRPHVMEPTP	921
Db	2360	PTTTTTPVTPRPTPTG--TQTP--TTPPTTTTTPVTP	2392
RESULT	2		
SLIP1	SLIP1_CLOTM	STANDARD:	PRT: 1664 AA.
AC	006852:		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	01-FEB-1996 (Rel. 33, Last annotation update)		
DE	CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER PROTEIN 1).		
GN	OLPB.		
OS	Clostridium thermocellum.		
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
CC	Clostridium.		
OX	NCBI_TaxID=1515;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NCIB 10682;		
RX	MEDLINE=93209931; PubMed=8458932;		
RA	Fujino T., Begun P., Aubert J.-P.;		
RT	"Organization of a Clostridium thermocellum gene cluster encoding the		
RT	cellulosomal scaffolding protein CfpA and a protein possibly involved		
RT	in attachment of the cellulosome to the cell surface.";		
RL	J. Bacteriol. 175:1891-1899(1993).		
CC	-1 SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.		
CC	-1 SUBCELLULAR LOCATION: CELL WALL.		
CC	-1 SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration -		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/sib.ch">http://www.isb-sib.ch/announce/sib.ch</a> ).		
CC	or send an email to <a href="mailto:license@sib.ch">license@sib.ch</a> ).		
CC	-----		
CC	EMBL: X67506; CAA47841.1; -		
CC	InterPro: IPR001119; SLH.		
CC	Pfam: PF00395; SLH; 3.		
CC	PROSITE: PS01072; SLH_DOMAIN; 2.		

FW	Signal	1	28	POTENTIAL
FT	CHAIN	29	1664	CELL SURFACE GLYCOPROTEIN 1.
FT	CHAIN	36	763	4 X 156 AA APPROXIMATE REPEATS.
FT	REPEAT	36	191	1.
FT	REPEAT	207	363	2.
FT	REPEAT	409	565	3.
FT	REPEAT	607	763	4.
FT	DOMAIN	771	1377	APPROXIMATE TANDEM REPEATS OF
FT	DOMAIN	1378	1449	T-P-S-D-E-P.
FT	DOMAIN	1453	1494	GLY/PRO/SER/THR-RICH.
FT	DOMAIN	1495	1565	SLH 1 (INCOMPLETE).
FT	DOMAIN	1566	1625	SLH 2.
FT	DOMAIN	1626	1646	SLH 3.
SO	SEQUENCE	1664	178194	SLH 4 (INCOMPLETE).
		AA:	MM:	5F396695BA9FE74B CRC64;
Query Match				
Best Local Similarity 17.4%; Score 942; DB 1; Length 1664;				
Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;				
QY	128	VLAAPT-P-KAETTTKGAALTTPEKEPTPTTPEKEPASTPTKEPTPTTIKSAPTTPEKAPATT	186	
DB	758	VVIDPAPAPKAASDEPIPDTPSDDEPTPS-----DEPTPS---DEPTPSDEPTPSD	804	
QY	187	TKSAPTTPEKAPATTPEKAPATTPEKPA-----PTTPEKAPATTTKSAPTTPEKAPATT	242	
DB	805	EPTSEPTPEEPPIPDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTSETEPEEPIPTDP	864	
QY	243	KPAETTPEKAPATTPEKPTPTTPEKAPATTPEKAPAT--TPKSPAPATAPKAPATTPEKAP	301	
DB	865	SDEPTPSDEPTPS---DEPTPS--DEPTP-SDEPTSETEPEEPIPDTPSDDEPTPSDEPTP	919	
QY	302	TTTPEKPA--PTTPEKPSPT--TPKEPAPTITKS--APTITKEAPATTTKSAPTTPEKPSPT	357	
DB	920	SDEPTPSDEPTPSDEPTSETEPEEPIPDTPSDDEPTPSDEPTPS-----DEPTP	968	
QY	358	TKPEAPTTPEKAPATTPEKAPATTPEKAPAT--TPKEPATTTKKAPATPEKAPATTPEKE	416	
DB	969	SDEPTPS--DEPTPS---DEPTPSDEPTSETEPEEPIPDTPSDEPTPSDEPTPSD---DEPTPS	1019	
QY	417	TAPTPTTKLPTTPEKLAAPTPEKRAPATTPEELAPTPEEPTPT--TPKEPAPTTPKAAAP	475	
DB	1020	-----EPTPSDE--PTPSDEPTPSD---EPTPSDEPTSETEPEEPIPDTPSDDEPT	1064	
QY	476	NTPEKAPATTPEKPA--PTPEKAPAPT--TPKEPATTTKGAAPTPEKAPATTPEKAPAPK	532	
DB	1065	TPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTSETEPEEPIPDTPSDEPTPSDEPTPS--DEPTPS	1122	
QY	533	ELAPTPTKEPT--STTSOKPAPTPEKGTAPTTPEKAPATTPEKAPATTPEKGTAPTTLKEPA	591	
DB	1123	D-EPTPSDEPTSETEPEEPIPDTPSDDEPTPSDEPTPS--DEPTPS---DEPTPSDEPT	1175	
QY	592	PT--TPPKAPAPKELAPATTTKGPTSTTSOKPAPT---TPKEPATTPEKAPAT--TPKKAPAT	646	
DB	1176	PSETPSEEP-----IPTDPSDEPTPSDEPTPSDEPTPSD--EPTPSDEPTSETEPEEPIPT	1229	
QY	647	TPETPTPTTSEVSTPTTKEPTTIHKSDEESTPELSAETPAKALENSEKEBEGVPTTKPA	706	
DB	1230	DTPSDEPTPSD--EPTPSDEPT---PSDEPTP--SDEPTP---SETEPEE--IPTDPSD	1277	
QY	707	ATKEPMTTAAOKTERLRTTPETTTAAAPKMTETATTTKTTESKITATTTQVTSITT	766	
DB	1278	EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDPTPSDEPTSETEPEEPIPTDTPSDEPTPS	1337	
QY	767	QDTPPTKITTTKTITTLAPKVTTTKTTITTEIMKKPEETAKPKRATYSKATTTPKQ--KP	825	
DB	1338	DEPTPSDEPT-----PSDEPTPSDEPTPSDEPTPSDEPTP	1372	
QY	826	TKAPAKPTSTKKPKTPMPVRKKPTTTPPKMTS-----TMDLNPJSRIA	870	
DB	1373	SETEPEEPTPTTTP-----TPPTSTPTTSGSGSGSGGGGGGGGTVPTSGPTPTPTPS---TP	1424	



DB 1123 GGTGTTTATTTTAVTTTAVTTTSSSTGTNSA-GKTTTGTTKSVPT-----YVTTL 1175  
 QY 975 SPFSPSPARRITEWVGIPSIDYFTRC-----NCEGRT 1009  
 DB 1176 APSAPVTPATN-----AVPTTITT--TECSATNNAGET 1207

RESULT 4  
 MUC1\_HUMAN STANDARD: PRT: 1255 AA  
 ID P15941: P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;  
 AC 01-UN-1990 (Rel. 13, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)  
 DE (EPISIALIN) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)  
 DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-  
 DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN  
 DE DF3).  
 GN MUC1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=90368716; PubMed=2394722;  
 RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;  
 RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";  
 RL J. Biol. Chem. 265:15294-15299(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90202794; PubMed=2318825;  
 RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;  
 RT "Episialin, a carcinoma-associated mucin, is generated by a  
 RT polymorphic gene encoding splice variants with alternative amino  
 RT termini.";  
 RL J. Biol. Chem. 265:5573-5578(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=90368715; PubMed=1697589;  
 RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,  
 RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;  
 RT "Molecular cloning and expression of human tumor-associated  
 RT polymorphic epithelial mucin.";  
 RL J. Biol. Chem. 265:15286-15293(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=91097524; PubMed=2268309;  
 RA Lancaster C.A., Peat N., Duhig T., Wilson D.,  
 RA Taylor-Papadimitriou J., Gendler S.J.;  
 RT "Structure and expression of the human polymorphic epithelial mucin  
 RT gene: an expressed VNTR unit.";  
 RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=90276413; PubMed=2351132;  
 RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,  
 RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,  
 RA Keydar I.;  
 RT "Human epithelial tumor antigen cDNA sequences. Differential splicing  
 RT may generate multiple protein forms.";  
 RL Eur. J. Biochem. 189:463-473(1990).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=90276414; PubMed=2112460;  
 RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,  
 RA Zrlhan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;

RT "A transcribed gene, containing a variable number of tandem repeats,  
 RT codes for a human epithelial tumor antigen. cDNA cloning, expression  
 RT of the transcribed gene and over-expression in breast cancer  
 RT tissue.";  
 RL Eur. J. Biochem. 189:475-486(1990).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91033045; PubMed=1688329;  
 RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,  
 RA Jeltsch J.M., Gartner J.M., Lathe R., Keydar I., Wreschner D.H.;  
 RT "Isolation and characterization of an expressed hypervariable gene  
 RT coding for a breast-cancer-associated antigen.";  
 RL Gene 93:313-318(1990).  
 RN [8]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE=88330762; PubMed=3417635;  
 RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,  
 RA Burchell J.;  
 RT "A highly immunogenic region of a human polymorphic epithelial mucin  
 RT expressed by carcinomas is made up of tandem repeats.";  
 RL J. Biol. Chem. 263:12820-12823(1988).  
 RN [9]  
 RP SEQUENCE OF 1-169 FROM N.A.  
 RX MEDLINE=90088473; PubMed=2597151;  
 RA Abe M., Siddiqui J., Kufe D.;  
 RT "Sequence analysis of the 5' region of the human DF3 breast  
 RT carcinoma-associated antigen gene.";  
 RL Biochem. Biophys. Res. Commun. 165:644-649(1989).  
 RN [10]  
 RP SEQUENCE OF 1-109 FROM N.A.  
 RC TISSUE=Thyroid;  
 RX MEDLINE=96183746; PubMed=8608966;  
 RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;  
 RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse  
 RT transcriptase polymerase chain reaction of the MUC1 gene.";  
 RL Int. J. Cancer 66:55-59(1996).  
 RN [11]  
 RP SEQUENCE OF 1-89 FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=96181716; PubMed=8604237;  
 RA Yu C.-J., Yang P.C., Shew J.-Y., Hong T.-M., Yang S.-C., Lee Y.-C.,  
 RA Lee L.-N., Luh K.-T., Wu C.-W.;  
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and  
 RT tissues.";  
 RL Oncology 53:118-126(1996).  
 RN [12]  
 RP SEQUENCE OF 1-46 FROM N.A.  
 RC TISSUE=Breast carcinoma;  
 RA Bulwela L., Liu Q., Lugmani Y.A., Gonn J.J., Coombes R.C.;  
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
 CC CYTOSKELETON.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM  
 CC IS ALSO PRODUCED.  
 CC -1- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING.  
 CC -1- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL  
 CC TUMORS, SUCH AS BREAST CANCER.  
 CC -1- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND STALL  
 CC ACID).  
 CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT  
 CC VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE  
 CC MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.  
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

DR EMBL: J05582; AAA60019.1; -  
 DR EMBL: M32738; AAA35804.1; -  
 DR EMBL: M32739; AAA35806.1; -  
 DR EMBL: J05581; AAA59876.1; -  
 DR EMBL: M61170; AAB53150.1; -  
 DR EMBL: X52229; CA36478.1; ALT\_SEQ.  
 DR EMBL: X52228; CA36477.1; ALT\_SEQ.  
 DR EMBL: M35093; AAB59612.1; ALT\_SEQ.  
 DR EMBL: Z17324; CA78972.1; -  
 DR EMBL: Z17325; CA78973.1; -  
 DR EMBL: M31823; AAA55757.1; -  
 DR EMBL: S81781; AAD14376.1; ALT\_INIT.  
 DR EMBL: S81736; AAD14369.1; ALT\_INIT.  
 DR EMBL: M21868; AAA36874.1; ALT\_SEQ.  
 DR PIR: A35175; A35175.  
 DR PIR: B35175; B35175.  
 DR PIR: S10218; S10218.  
 DR GlycoSuiteDB: P15941; -  
 DR MIM: 158340; -  
 DR MIM: 113720; -  
 DR InterPro: IPR000082; SEA.  
 DR Pfam: PF01390; SEA; 1.  
 DR SMART: SM00200; SEA; 1.  
 DR PROSITE: PS50024; SEA; 1.  
 KM Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  
 Repeat; Alternative splicing.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 1255 MUCIN 1.  
 FT DOMAIN 24 1162 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1163 1186 POTENTIAL.  
 FT DOMAIN 1187 1255 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 81 960 44 X 20 AA TANDDEM REPEATS.  
 FT DOMAIN 1034 1151 SEA.  
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 975 975 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1133 1133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VASPLIC 19 19 T-> TATTAKPRAT (IN ISOFORM B).  
 FT VASPLIC 20 22 MISSING (IN ISOFORM C).  
 FT VASPLIC 20 31 MISSING (IN ISOFORM D).  
 FT VASPLIC 126 905 MISSING (IN POLYMORPHIC EPITHELIAL ISOFORM).  
 FT VASPLIC 1077 1087 FLOIRKGGFL -> VSLGSPMLP (IN SECRETED ISOFORM).  
 FT VASPLIC 1088 1255 MISSING (IN SECRETED ISOFORM).  
 FT CONFLICT 2 2 T-> A (IN REF. 11).  
 FT CONFLICT 134 134 P-> Q (IN REF. 9).  
 FT CONFLICT 154 154 P-> Q (IN REF. 9).  
 FT CONFLICT 1021 1021 S-> T (IN REF. 3).  
 FT CONFLICT 1251 1251 A-> T (IN REF. 3).  
 SQ SEQUENCE 1255 AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;

Query Match 12.0%; Score 651; DB 1; Length 1255;  
 Best Local Similarity 27.8%; Pred. No. 1.3e-19;  
 Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;

DB 75 TSLVNNKETTETTKETTNNKOTSDGKEKTSAKTOSIEKTSKAD-LAPTSKIYLAKEP 133  
 16 TFLTV-----VTGSHASSTPGGEKETSATORSSVPSSTENNAVSMSSVLSHS 65  
 QY 134 P-KAETTKGP--ALTTREP-----TPTPKREPASTTPEPTTIT--K 173  
 DB 66 PGGSGSTTGODVTLAPATEPAGSAATWGODVTSPTVTRPALGSTITPAHVDTSAPDK 125  
 QY 174 SAPTPKEAPATTTSAPTPKEAPATTTPK-----APTPKEAPATTTPKEAPATT 227  
 DB 126 PABGSTAPRAHGVTSAPDT--RPAGSTAPRAHGVTSAPDT--RPAGSTAPRAHGV 179  
 QY 228 SAPTPKEAPATTTP-----KKPAPTTPREP-----APTTPKEPTTP 265  
 DB 180 SAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGV 239

QY 266 ----KEAPATTKP-----APTTPKEAPATPK----KPAPTTPKEAPATTPK----E 306  
 DB 240 SAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGV 299  
 QY 307 PAPTPPKESPPTTPKEAPATTP-----TKSAPTTTKEAPATTPKSAPTTPKEPSPTTKP-- 361  
 DB 300 SAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGV 356  
 QY 362 ----APTTPKEAPATTPK-----KPAPTTPKEAPATTPK-----EAPTTTKPAPTAPKEP 409  
 DB 357 GVTSAADTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 416  
 QY 410 APTTPKEAPATTPPKLTPTTPEKLAAPTTPEKAPATTPPELAPTTPEEPPTTPEEAPATT 469  
 DB 417 GVTSAADTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 472  
 QY 470 PKA-----AAPNTPKEAPATTPK-----EPAPTTPKEAPATTPKETAPTTPKGTAPTTPKEP 521  
 DB 473 PRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 528  
 QY 522 APTTP-----KKPAPKEAPATTPKEPTSTTSKAPAP--TTPKGTAPTTPKEAPATTP 569  
 DB 529 GSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 583  
 QY 570 TPKEAPATTPK-----TAPTTKEAPATTP-----KKPAPKEAPATTPKGTAPTTPKEP 615  
 DB 584 TRPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 642  
 QY 616 SDKPAP--TTPK-----ETAPTTPKEAPATTP-----KKPAPTTPETPTPTTSEVS 659  
 DB 643 DTRPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 700  
 QY 660 TPTTPKEPT-----IHKSPDESTPELSAPPTKALENSKEGVPVTTKPAATKEM 712  
 DB 701 ADTPRAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 753  
 QY 713 TTTAKDTERLRTTPETT-----TAAPKMTKETATTEKTESKITATTTGVTSTT 765  
 DB 754 --PANGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 807  
 QY 766 TDDTTP--FKITTLKTTTLAPKVTITTKITTT--EIMNKPETAKPRADRANSATTPK 821  
 DB 808 PGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAH 867  
 QY 822 POKPTKAPKPTSTKKPKTPMVRKPTTPPKMTSTJMEPELNPSTRIAEAMLQTTTBN 881  
 DB 868 P--GSTAPRAHGVTSAPDTRP--APGSTAPRAHGVTSAPDTRP-----PG 909  
 QY 882 QT--PNSKIVEVNPKESDAGAGETPRHMLLRPHVMEVTPDMDYLPRVNOGIIINPM 940  
 DB 910 STAPRAHGVTSAPDTRPAPAGSTAPRAH-----GVTSAADNRPALGSTA--PPV 955  
 QY 941 SPEFTICNGRPVDGLTTLRNGTLVAFRGHYFMMLSPSPSPS 981  
 DB 956 HVTSAAGSASASATLVHNGTSARATTTTAPASKSTPFSIPS 996

RESULT 5  
 FPL\_MYTED STANDARD; PRT; 875 AA.  
 AC Q25460;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADHESIVE PLAQUE MATRIX PROTEIN (POLYHENOLOGIC ADHESIVE PROTEIN) (FOOT PROTEIN 1) (MEP1) (FRAGMENT).  
 GN FPL.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea; Mytiloidae; Mytilidae; Mytilus.  
 NC NCBI\_TaxID=6550;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA MEDLINE-91025829; PubMed-1367451;  
 RA Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;  
 RT "Structural and functional repetition in a marine mussel adhesive  
 RT protein.";  
 RL Biotechnol. Prog. 6:171-177(1990).  
 RN [2]  
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.  
 RX MEDLINE-83135732; PubMed-6298211;  
 RA Waite J.H.;  
 RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and  
 RT hydroxyproline-containing decapeptide in the adhesive protein of the  
 RT mussel, Mytilus edulis L.";  
 RL J. Biol. Chem. 258:2911-2915(1983).  
 CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
 CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
 CC -1- PPM: THE DECAPEPTIDE A-K-P-S-Y-P-T-Y-K IS POST-TRANSLATIONALLY  
 CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE  
 CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE  
 CC (DOPA) DERIVED FROM TYROSINE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X54422; CAA38294.1; -  
 DR Interpro: IPR002964; Adhesive\_plaq.  
 DR Interpro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PRO1216; ADHESIVE1.  
 DR PRINTS: PRO1217; PRICHEXTENS.  
 KW Repeat; Hydroxylation.  
 FT NON\_TER 1  
 FT DOMAIN 67 870  
 FT SEQUENCE 875 AA: 100412 MW: 66A85312748CAACE CRC64;  
 SQ  
 Query Match 10.3%; Score 556.5; DB 1; Length 875;  
 Best Local Similarity 28.7%; Pred. No. 5,2e-16;  
 Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;  
 QY 42 TONKAVSTSPKITTAKPINP-----RP-SLPPNSDTSKETSLTVNKETTVET 87  
 DB 1 TKHBPVYKPVYTSABYKPPYQPLKKKVDYRPTKSYPPYG-SKTVYLPKLKLSYK 59  
 QY 88 KETTTKNGSTDC--KEKT---SAKETOSIEKTSKDLAPTSKVLAKPPKAEITTKG 142  
 DB 60 PIKTTYNAKINPPVYKPKMYPPYKPKPSYPPYKSKPYKPKITTYPPYKAKPSY- 117  
 QY 143 PALTPKPEPTTPPEK-----PASTTPKEPTPTTIKSAPTTP-----KEPAPTTKSAPTT 193  
 DB 118 PSSYKPKKTYPPYKPKLTYPPYKPKPSYPPYKPKPSYPSKTKTYSSTKAKPSY 177  
 QY 194 PKEPAPTTKEPA--PTTPKEPA-----PTTTKEPA--PTTTKSAPTTPK--EPAPTT 240  
 DB 178 P-----PTYKAKPSYPPYKAKPSYPPYKAKPYKAKPTYSYKAKPSYPPYKAKPTY 233  
 QY 241 PKKPA--PTTPKEPA--PTTPKEPT--PTTPKEP--APTKEAPPTTPKPA--PTAPK 289  
 DB 234 KAKSYPPYKAKPSYPPYKAKPSYPPYKAKPYKAKPYKAKPYKAKPSYPPYKAK 293  
 QY 290 KPA--PTTPKEPA--PTTPKEPS--PTTPKEPA--PTTPKEPA--PTTPKEPA--PTTPKEPA- 339  
 DB 294 KPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPYKAKPTY 353

QY 340 PTTTSAPPTKEPSPPTTTKEPAPTTTPKPA--PTTPKEPA--PTTPKEPA- 395  
 DB 354 PSTYKAKPSYPPYKAKPSYPPYKAKPYKAKPSYPPYKAKPSYPPYKAKPSY 409  
 QY 396 TTTKRPAPTAPEPAPPTTPKETAPTTPPKLPTPTPEKLAPTTPPEKAPPTPELAPTTPE 455  
 DB 410 PPTYKAKPYKAKP-----TYPST-YKAKPSY-----PSIKAKPSYPPYKAKPYKAK 457  
 QY 456 EPTPTPEEPAPTTKKAAPNTPKEPAPTTTPKEPAPTTPE-----PAPTTPEPAPT 509  
 DB 458 KPT-----YPSITYKA-----KPSYASAKAKPSYPPYKSKSPSSYKPKKTYPPYK 506  
 QY 510 PKGAPTLTKEPAPTTPKKPAKELAPTTTKPTSTSDKRAPPTPKGAPT-----PKE 565  
 DB 507 PKLITKPPYK-PKPSYPPSYKPKTTPPYK-----PKISYPPYKAKPSY 551  
 QY 566 PAPTPKPEPAPTTPKGT--APTTPKEPA--PTTPKPA--PKELAPTTTKG-PTSTSD 617  
 DB 552 PATYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYKAKPYKAKPSYPPYK 611  
 QY 618 KPAPTTPEETAPTTTPKEPA--PTTPKPA-PTT-----PETPPTTSEVSTPTTPKE--- 666  
 DB 612 KAKPSYPP--PTYKAKPSYPPYKAKPYKAKPSYPPYKAKPSYPPYKAKPSY 667  
 QY 667 PTTIKSPDESPBELSAPPTPKALENSPK-----EBGVPTT--KTPA-----ATKPEMTT 715  
 DB 668 PPT-YKAPPSYPPYKAKPYKAKPTNPSTYKAKPSYPPYKAKPSYPPYKAKPSYPT 726  
 QY 716 AKDKTTERDLRTPTETTTAAPKMTKETATTEKTESKITATTTQVSTTQDDTTPPKIT 775  
 DB 727 YKAKPTYKAKPYKPSYKAKP--TYKAKPTYPYKAKP-----PSYPPYKPKPSYPP 777  
 QY 776 TLKTTTLAPKVTYTKKTTTTEIMKPEETAKPKRAATNSKATTTPKPKPTKAPK---P 832  
 DB 778 TYKSSISYPSYKPKKTYPPPT--YKPKLTYPPYK-----PKPSYPSYKPKKITYPP 826  
 QY 833 TSTKPKTMPRVKRPKPTTPP-----RKMTSTMPELNPTSR 868  
 DB 827 STYKAKPSYPPYKSKTSYPPYKKNKISYPSYKAKTISYPAKPKPTNR 874  
 RESULT 6  
 ZAN\_HUMAN STANDARD; PRT; 2700 AA.  
 AC 09Y493; 000218;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ZONADHESIN (FRAGMENT).  
 GN ZAN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 1-2379 FROM N.A.  
 RX MEDLINE-99018118; PubMed-9799793;  
 RA Glockner G., Scherer S., Schatteroy R., Botright A., Weber J.,  
 RA Tsai L.C., Kosenthal A.;  
 RT "Large-scale sequencing of two regions in human chromosome 7q22:  
 RT analysis of 650 kb of genomic sequence around the Epo and CYP11 loci  
 RT reveals 17 genes.";  
 RL Genome Res. 8:1060-1073(1998).  
 RN [2]  
 RP SEQUENCE OF 2338-2700 FROM N.A.  
 RX TISSUE-TESTIS;  
 RX MEDLINE-97271566; PubMed-9126492;  
 RA Gao Z., Harumi T., Garbers D.L.;  
 RT "Chromosome localization of the mouse zonadhesin gene and the human  
 RT zonadhesin gene (ZAN).";  
 RL Genomics 41:119-122(1997).  
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA

CC		-I- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC		-I- SUGCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC		-I- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC		-I- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE ZONA PELLUCIDA.
CC		-I- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS, THE MUCLN-LIKE DOMAIN MIGHT INHIBIT IMPROPERADTAE TRAPPING OF SPERMATOZOEA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC		-I- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIAE COVALENT OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCLN MCC2).
CC		-I- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC		-I- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
-----		
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between The Swiss Institute of Bioinformatics and the EMBL Outstation - CC between The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by end user for commercial purposes requires a license agreement (See http://www.isdb.ch/announce/or send an email to licensee@isdb.ch).
CC		
DR	EMBL, AF05356;	AAC7890.1; ..
DR	EMBL, U83191;	AAc51208.1; ..
DR	MIM, 602372;	--
DR	InterPro:	IPIR000561; EGF-like.
DR	InterPro:	IPIR000998; MAM.
DR	InterPro:	IPIR002965; P-rich_extensn.
DR	InterPro:	IPIR002919; TIM.
DR	InterPro:	IPIR003328; TLTA.
DR	InterPro:	IPIR001007; VWC.
DR	Pfam:	PF00629; MAM_3.
DR	Pfam:	PF01825; TTL; 4.
DR	Pfam:	PF02345; TTL; 4.
DR	Pfam:	PF00094; vwd; 4.
DR	PRINTS,	PR01217; PRICEXTENSNS.
DR	SMART;	SMD0137; MAM; 2.
DR	SMART;	SMD0114; WVC; 1.
DR	SMART;	SMD0011; VWC_def; 3.
DR	SMART;	SMD00216; VND; 3.
DR	PROSITE,	PS01186; EGF_2; 3.
DR	PROSITE,	PS00740; MAM_1; 1.
KW	Glycoprotein;	Transmembrane; Cell adhesion; Repeat.
FT	NON_TER	1
FT	DNAIN	<1
FT	DMAIN	112
FT	DMAIN	161
FT	DMAIN	322
FT	DMAIN	483
FT	DMAIN	953
FT	DMAIN	1066
FT	DMAIN	1455
FT	DMAIN	1862
FT	DMAIN	2292
FT	DMAIN	2293
FT	DMAIN	2684
FT	CARBHYD	74
FT	CARBHYD	403
FT	CARBHYD	1023
FT	CARBHYD	1099
FT	CARBHYD	1618
FT	CARBHYD	1737
FT	CARBHYD	1832
FT	CARBHYD	1878
FT	CARBHYD	2136
FT	CARBHYD	2505
FT	CARBHYD	2379
FT	NON_TER	2700
SEQUENCE	2700 AA;	293013 MW; 80E6COCB12277B1 CRC64;

Query Match	10.3%	Score 555.5	DB 1.5	Length 2700
Best Local Similarity	32.6%	Pred. No. 1.4e-15		
Matches 199	Conservative 64	Mismatches 240	Indels 107	Gaps 32

  

QY	58	PINRPSLPNPSDTSKETSLSLVNKKETTVETKETTNTKOTSIDGKEKTTSAKETSISIEKT	117
Db	445	PKVLPVLEIPVSPVSS-----TGPSEETGLTENDPTISTK-----KPTVSIENK	487
QY	118	SAKDAPISKVLAKPTKAEETTKGAPLALTPPEPTPTPKKEPASTPKPEPT-----PTPI	172
Db	488	SVTTEKPP-----VPEKEPTLTPEKPTISTEK---PTIPSEKPNMSEKPTIPSEKPTIL	539
QY	173	KSAPITPEKPAPTTKSAPATTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPA---PTTKSAPT	231
Db	540	TEKPTIPSE-KPTIPSEKPTISTEKPTVPTSE--PTTPEETTTTMEEEVATPEKPSIPT	596
QY	232	TPKEPAPTTPKKPAATTTKKEPAPTTKKEPTPTTKKEPAPTTKKEPAPTTKKEPAPAPKAP	291
Db	597	--EKISITPEK---PTISMEETITSTEKPTICPEKPTIPTEK---PTITSEKSTISPEK-	647
QY	292	APTPKEPAPTTPKKPAATTTKKEPAPTTKKEPAPTTKKEPAPTTKSAPTTKKEPAPTTKSAPT	351
Db	648	PTTTE-KPTIPTEKPTISTEKPTIPPEK-PTISEKPTIPTEKLTPT---EKPTIPT	701
QY	352	EPSPTTKEPAPTTKKEPAPTTPKKPA-----PTTPKEPAPTTKKEPAPTTTKKRAPAP	406
Db	702	EKPTISTEE-PTTTEETITSTEKPSIPEMEKPTLTPEETTTTSVEETTTISTEKLTPM--	757
QY	407	KEPAPTTKEPAPT-----TPKKLTPTPEKLA-----PTPEKPAPTTPEELAPTPPEE	456
Db	758	EKPTISTEKPTIPTEKPTISPEKLTIPPT-EKLTIPTEKLTIPTEETISTEK--TIPPE	814
QY	457	PTTPPEEPAPTTPKAAPTNPKEPAPTTPKKRAPPTPKKRAPPTPKKRAPTTPKKPTAPT	516
Db	815	KPTISEKPTISTEK-----PTTPE-KPTIPTE-----ETTISTEKLTPPT	855
QY	517	TLKEPAPTTPKKAPKELAPTTTKEPTST-----SDKPAPTPKGAPTPKKEPAPTT	570
Db	856	--EKPTISEKLTPTTEKPTISTEKPTIPTEKLTIPTEKPTIPT--EKLTAIR	911
QY	571	PKPAPTTPKGTAPTTLKEPAPTTKPKKPAKELAPTTTKGPTSTSDKAPTTPKET--A	628
Db	912	PNHPSPTA-TGIALVMSPHASPTPTVS---ILGTTTSSRSTGMSCP-PNARYESCAC	966
QY	629	PTTPKEPAPT	638
Db	967	PASCKSPRS	976

  

RESULT_7		
EXTN_TOBAC	STANDARD	PRT: 620 AA.
AC	PI3983;	
DT	01-JAN-1990 (Rel. 13, Created)	
DT	01-JAN-1990 (Rel. 13, Last sequence update)	
DT	01-MAR-1992 (Rel. 21, Last annotation update)	
DE	EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).	
CN	HROGNTN	
OS	Nicotiana tabacum (Common tobacco).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
OC	Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.	
OX	NCBI_TaxID=4097;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. XANTHI; TISSUE=Leaf.	
RX	MEDLINE=90128263; PubMed=2612909;	
RA	Keller B., Lamb C.J.;	
RT	"Specific expression of a novel cell wall hydroxyproline-rich	
RT	glycoprotein gene in lateral root initiation.";	
RL	Genes Dev. 3:1639-1646(1989).	



```

CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC MAIN ROOT.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X13885; CAA32090.1; -.
CC PIR: S06733; S06733.
CC Repeat: Cell wall; Glycoprotein; Signal; Structural protein;
CC Hydroxylation.
CC CHAIN 1 2
CC REPEAT 70 73 EXTENSIN.
CC REPEAT 148 151 H-A-P-P.
CC DOMAIN 229 242 H-A-P-P.
CC REPEAT 229 235 2 x 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
CC REPEAT 236 242 1.
CC DOMAIN 205 620 2.
CC FT 499 600 3 x APPROXIMATE TANDEM REPEATS.
CC SEQUENCE 620 AA; 65406 MW; 641DD2278AB26524 CRC64;
SO
Query Match 10.2%; Score 551; DB 1; Length 620;
Best Local Similarity 27.2%; Pred. No. 6,6e-16;
Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;
CC
CC 136 AETTK-GPALTTP--KEPRTT---PKEPASTPKKE--PPTTKASAPTPKKEAP 184
CC 24 AEATTOVGGLYPVTSQPPSSIGLSPASAPTTTPSRGHVPS--RHAPRHAYAP 80
CC 185 TTTSAPTPKREP-----APTTKKEAPPT---PKKEAPTTTKEAPTTTTSATTTKE 235
CC 81 SHGLPSPVGGPRRHGHLPRSGRNPSPSVISPSHPPSYGAPRSHGCHLPSHGQR 140
CC 236 PAPTTPKKEAPTT--TKKEAPTTKEPTTPPKKEAPTTK--EPAPTTKEAPATAKK 290
CC 141 P-----PSSHGHAHRRSGGHTTPRRGCHNRRSPSRKGNHRRPTTYAQRPPTYIS 193
CC 291 PAPTTPKKEAPTTKKEAPTTTKEPSPTPK--EPAPTTKSAPTTKEAPTTTTS-- 345
CC 194 PSPGV--QRPPT--YSPRRTHVQTPRSPRGNQORPRTGHRNHRPTGHRNHRPTGHR 249
CC 346 ---APTPKESPTTKEAPATTTKEAPTTKKEAPTTKKEAPTT--TKKEAPTTTK 399
CC 250 RHLRPSRKRQOPRTYSPRPVAYQSPRTYSPRPVAYSPRPVAYSPRPVAYSP 309
CC 400 KPAAT---APKERPAPTTKKEAPTTKPKL--TPTT-----EKLAPTTKEKAPATPEL 449
CC 310 PPTTPPTSPRPVAYSPRPVAYSPRPVAYSPRPVAYSPRPVAYSPRPVAYSP 369
CC 450 APTPKEEPTTPTEAPATTKKAAP---NTPKKEAPTTKKEAPATTKKEAPTTKETA 506
CC 370 PPPPSSPPSPSPSPPTTYEQSPRPVAYSPRLAPRTYSPPTT--YSPRPVAYAPR 427
CC 507 PTTKGAATTLKKEAPTTKKEAPTTKKEAPTTTKEPTSTTDKAPATTTPKGAATTPKEP 566
CC 428 P-----LPRTYSPRPVAYSPRPVAYSPRPVAYSPRPVAYSPRPVAYSP 481
CC 567 APTPKEAPATTTPKGAATTLKKEAPATTKKEAPATTKGAATTTGKTSTTSKAPATTKE 626
CC 482 PPPSPISPPRPVAYSPRPVAYSPRPVAYSPRPVAYSPRPVAYSPRPVAYSP 538
CC 627 TAPTTPKEAPTT--TPKKAPATTPEPTTPSTSEVSTPTTKEPTTIHNSPDESTPELSAE 684

```

```

DB 539 PPPRQINSPPRRHMQPRTTPYGGPSPRPSPRPSPRPSPRPSPRPSPRPSPRP 598
CC 685 PPKKALENSPKRGVPTTKTPATPK 710
CC 599 PSP-----PTTYSPPSPPP 612
CC -----
CC RESULT 8
CC ID NEFH_MOUSE STANDARD: PRT: 1087 AA.
CC AC P19246; Q61959;
CC DT 01-NOV-1990 (Rel. 16, Created)
CC DT 01-AUG-1992 (Rel. 23, Last sequence update)
CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE NEUROFILAMENT TRIPLETT H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
CC DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
CC GN NEFH OR NEFH.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC OX NCBI_TaxID=10090;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=89121513; PubMed=3220257;
CC RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosfeld F.,
CC RA Mushynski W.;
CC RT "Sequence and structure of the mouse gene coding for the largest
CC RT neurofilament subunit."
CC RL Gene 68:307-314(1988).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=89089138; PubMed=3145094;
CC RA Sheidman P.S., Carden M.J., Lees J.F., Lazarini R.A.;
CC RT "The structure of the largest murine neurofilament protein (NF-H) as
CC RT revealed by cDNA and genomic sequences."
CC RL Brain Res. 464:217-231(1988).
CC RN [3]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
CC RA Carden M.J.;
CC RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC CC NEFH HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NEFH IS
CC CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC CC THOUGHT THAT PHOSPHORYLATION OF NEFH RESULTS IN THE FORMATION OF
CC CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC CC OF AXONAL CALIBER.
CC CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M24496; AAA39813.1; -.
CC DR EMBL: M23349; AAA39813.1; JOINED.
CC DR EMBL: M24494; AAA39813.1; JOINED.
CC DR EMBL: M24495; AAA39813.1; JOINED.
CC DR EMBL: M35131; AAA39809.1; ALT_FRAME.
CC EMBL: Z31012; CAA83229.1; -.

```





RA Inoue K., Takeuchi Y., Takeyama S., Iamada E., Tamada T., Ose N.  
RA Hareyama S.:  
RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and  
RT its evolutionary implications";  
RL J. Mol. Evol. 43:348-356(1996).  
CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
CC -1- PGM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND  
CC ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D63777; BAA09850.1;  
DR InterPro: IPR002964; Adhesive\_p1qg.  
DR InterPro: IPR002965; P\_rich\_extensn.  
DR PRINTS: PR01217; PRICHCXTENSN.  
DK SIGNAL: Repeat; Hydroxylation.  
FT SIGNAL: 1 20 POTENTIAL.  
FT CHAIN: 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.  
FT DOMAIN: 21 41 NONREPETITIVE LINKER.  
FT DOMAIN: 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-P-[ST]-[ST].  
FT DOMAIN: 184 192 NONAPEPTIDE 1.  
FT DOMAIN: 213 221 NONAPEPTIDE 2.  
SQ SEQUENCE 872 AA; 10167 MW; 98CCT0DD7C5FF3C4 CRC64;

Query Match 9.6%; Score 518.5; DB 1; Length 872;  
Best local similarity 29.2%; Pred. No. 1,7e-14;  
Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54

OY 19 VDEAGSGLDNGEPKVTTPDTST--NQH-----KVNSTSPKITAKPINPRLSPNS 69  
DB 29 VYGSAYVSGASGAVK-TLPESHGYGSKHVNYVPMMKIRP-PYI--SKMSYPAPYPKGV 84  
OY 70 DTSEKSLTVNKKETVEYETKEETTNNQTSNDGKEKTTSKAKETOSIEKTSKADLAIPSKVL 129  
DB 85 YPKRRVYPTGSKNVNPPIPKPLAKKLSSYKAIKITVPAVKANTSVPPSYK----HKIT 139  
OY 130 AKPTPKAETTTKGPAITTPKEPT-PTTPKPBPASTTPKEPTTIKSAPTTPKBPATTTK 188  
DB 140 YPPPYK-----PKRIYV--PYTKOKPSYPSYSKPKTTTTPPIYK-----PKITIYPTTK 185  
OY 189 SAPV-TPKEBAPT--TTKEBAPTTPKEBAPTTPKBPATTTKSAPTTPKBAPT--PKK 243  
DB 186 RKBSYTYPKPKATVPPYTKKIYV-----PLYKKRPSTV-----PYRKPTTYPPYTKPKI 236  
OY 244 PAPTPKBPAPT---PKERTPTT--PKBPATTPKBPATTPKBPAPT-APPKBPATTP 296  
DB 237 SYPSIYRKASYVSYSKSKTTPYTKKISYPPYKRKPSYV---PTYKPKATTPPT 292  
OY 297 KEPATTPKBPATTTTKSPSTTPKBPAPTTPKSAPT-----TTKEBAPTTTKSAPTTK 351  
DB 293 K-PKRSYPTTKPKITTYPTT-YKKRPSYPTKYKOKPSYPIYKSSYPTYSKSKTTPP 350  
OY 352 EPSPTTKEBA-PTTPKBPATTPKBP-APV-TRK-----EAPPTPKBPATTPTKKP 401  
DB 351 TYKPIKITVPTTYKPKPSYPSYKPKKITSYPTKKIIVPYTKPKPSYPSYKFKTITPP 410  
OY 402 APATAKEBAPTTPKKTAP-----TPPKKITPTTPKLAPTTPKBPAPTTPBELAFTPEE 456











```
Db 645 ---KAKEPPKVEEKEKTPATPKTEVYESKDEAPKEAOKP-KAEKEPLTEKPKDSPEGA 700
Oy 510 PKGTAPTTKEBPAPTTPKKPAKELAPTTTKEPTSTSDKPAPTTPKGTAPTTKBPAPT 569
Db 701 KKEEA---KEKKAAPPEETPAKLGVEEAKPKKAEADAKA-----KEPSKP 744
Oy 570 TPKEBAPTTPKGTAPTTTKEBPAPTTPKKPAKELAPTTTGTSTSTSDKPAPTTPKETAP 629
Db 745 SEKE---KPK-----KEEVPAPAEKKDKTE-----EKTTESKKKEEKKMEAK 784
Oy 630 TTPKEBAPTTPKKPAPTTPETPPPTTSVSTPTTKEPTTIHKSPDE 676
Db 785 AKED-----KGLPOEPSKPKTEKAESSTDOKDSOPSEKAPED 824
```

Search completed: April 26, 2002, 16:20:03  
Job time: 388 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:17:20 ; Search time 49.78 Seconds

(without alignments)  
457.932 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1212

Perfect score: 5416  
Sequence: 1 VKDNKRNRTKKRPPV.....SPIDVFTRCNCEGKTFPEK 1013

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCFUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	522	9.6	805	4	US-09-103-429A-4
2	508	9.4	1837	3	US-08-928-361B-5
3	506.5	9.4	744	6	5202236-25
4	498.5	9.2	786	4	US-09-103-429A-3
5	489	9.0	1721	3	US-08-700-651-5
6	489	9.0	1721	3	US-08-928-361B-6
7	488.5	9.0	826	1	US-07-638-431-2
8	488.5	9.0	826	1	PCF-US92-00018-2
9	476.5	8.8	652	6	5202236-13
10	452	8.3	960	4	US-09-219-849-5
11	424.5	7.8	1867	2	US-08-479-537A-5
12	424.5	7.8	1867	4	US-09-083-116-5
13	424.5	7.8	2035	2	US-08-479-537A-2
14	424.5	7.8	2035	4	US-09-083-116-2
15	419.5	7.7	2476	2	US-08-276-967-2
16	419	7.7	1185	4	US-09-041-886-23
17	417	7.7	829	1	US-08-642-255-132
18	417	7.7	829	1	US-08-397-633A-53
19	417	7.7	837	1	US-08-175-155-68
20	417	7.7	837	1	US-08-477-509B-103
21	417	7.7	837	1	US-08-642-255-101
22	417	7.7	837	2	US-08-707-237A-75
23	417	7.7	837	3	US-08-482-085B-103
24	417	7.7	897	1	US-08-397-633A-50
25	413.5	7.6	907	3	US-08-783-774-2
26	413.5	7.6	907	5	PCF-US95-04611A-19
27	404.5	7.5	408	1	US-07-609-716-65

28	404.5	7.5	408	4	US-08-475-411A-65	Sequence 65, App1
29	404.5	7.5	408	4	US-08-478-029A-65	Sequence 65, App1
30	398.5	7.4	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	7.4	682	1	US-08-397-633A-36	Sequence 36, App1
32	381	7.0	1848	4	US-08-296-791-6	Sequence 6, App1
33	381	7.0	1848	5	PCF-US95-10661A-6	Sequence 6, App1
34	376	6.9	360	4	US-09-219-849-6	Sequence 6, App1
35	374.5	6.9	1537	3	US-08-325-267A-2	Sequence 2, App1
36	373	6.9	1231	3	US-08-904-263A-4	Sequence 4, App1
37	371	6.9	761	2	US-08-707-237A-84	Sequence 84, App1
38	371	6.9	762	1	US-08-642-255-114	Sequence 114, App
39	371	6.9	762	1	US-08-397-633A-26	Sequence 26, App1
40	370.5	6.8	1064	1	US-08-642-255-62	Sequence 62, App1
41	368	6.8	762	1	US-08-320-559-28	Sequence 28, App1
42	367.5	6.8	1187	1	US-08-320-559-28	Sequence 31, App1
43	367.5	6.8	1187	3	US-08-545-860D-28	Sequence 28, App1
44	367.5	6.8	1187	5	PCF-US94-04496-28	Sequence 28, App1
45	367.5	6.8	1187	5	PCF-US94-04496-28	Sequence 28, App1

#### ALIGNMENTS

```
RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
;
GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Nuc1n
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Nuc1n
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Plunisi & Michaels, P. C.
; STREET: 118 No. 6187558ch T1oga
; CITY: Itasca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
;
INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
;
US-09-103-429A-4
;
Query Match 9.6%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 2,4e-26;
```

QY 125 TSKVLAKPTPKAEFTTKCPALTTKEPRTPTPKEDASTPEEPPTIKNSAP-----TTP 179  
Db 9 TALGVAAARPEVSDAEKNPALHEHPDXX--PAEOXLLPEVECTCFYVCENGKPIAP 66

QY 180 KEAPATTT-----TKSAPTTKEPATTTTKEPATTTKEPATTTTKEPATTTK 227  
Db 67 RCGACGTETEKFSAGCNVAHALAGCTLPGPPAETT---QAPATTQ--APTITQ--APTITT 119

QY 228 SAPPTPKEPAPTTPKKPAPTTPKKPAPTTKEPRTPTKEPATTTKEPATTTKEPATTA 287  
Db 120 QAPTTTTQAATTTT-----QAPTTQAATTT---QATTTQAATTTT----- 156

QY 288 PKKPAPTTKEPATTTKEPATTTKEPSPTTKEPATTTKSAPTTTKEPATTTKSAP 347  
Db 157 ---QAPTTTQ-APTITTO-APTITTO-----APTITQ-APTITTO-APTITIOAA 198

QY 348 TTPKEPSPTTKEPATTT--KEPATTTKKAPPTTKEPATTTK----- 391  
Db 199 TTPAATTPAAT--TPAATTPAATTPAATTPGVPAPS---APWBPICELLNCCPADFD 253

QY 392-----EBAPTTK 399  
Db 254 IHLILPHDKCYKLFTOCSNGYTFFQRCRCEGLXFNNYVGRDSPANVEDGELISAPRYTE 313

QY 400 K-----PA----- 402  
Db 314 GNEDEDIDIGDLLDNCGPANFEIDMLPHGNCRDXYOCVHGNIIVERCAGTHSEFLQ 373

QY 403-----FPAP--KEPAT-------TPKET 417  
Db 374 QCDDHIELVGCTLPGESEVEDDEDACTGCWYCPTERIEMERPLNCCPADFESIDLHPHES 433

QY 418-----APTTPKKLIPTT-----PEKLAPTTEKAP 443  
Db 434 DCGQYLQCVHGTIARPCGNLHFSPATOSCESPSYAGCOVECDSDMNCSTAAPAP 493

QY 444 TTPEELAPTTPEERTPTTBEAPTTPKKAAPVTPEAPTTKEPATTTKEPATTPK 503  
Db 494 TAAPTAAPTAPAASSTVVPPA--TPPATTAAPVPTTAIPT---PAPATAAPTAAPTTA 548

QY 504 ETAPTT---PKGTAPTTLKEPATTTKEPKAPKELAPTTTKEPTSTTSXKADPTPKGA 560  
Db 549 PESTIVTVPTNAAPA--APT---AVPEIPITVSAPTAAPT--AAPTAAPTAP 598

QY 561 TTPKEPATTTKEPATTTKGTAPTTLKEAPTTTKEPKAPKELAPTTTGPTST-TSQK 619  
Db 599 TTAAPLEIPT---VTSPTAAPT--AAAPMT-----TVTVPPTAAPTAA 641

QY 620 APTTPKEATTPKEPATTTPKKPAPTTPEPTTSEVSTPTTKEPTTIHKSPDESTP 679  
Db 642 AP-----NTTYTAAPTAPTT-AAPAPRTIVTVPTAAPTAAAP-----PIVAH-----AP 685

QY 680 ELISAEPKALENSPKEGVPTTKTPAIKPE 711  
Db 686 NTAAP-----VTTTSAPATTT-PE 703

RESULT 2  
US-08-928-361B-5  
Sequence 5, Application US/08928361B  
Patent No. 6071518  
GENERAL INFORMATION:  
APPLICANT: Petersen, Carolyn  
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS AND FRAGMENTS  
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
STREET: 365 Sherman Avenue, Suite 6

CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94306-1840  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/928,361B  
 FILING DATE: 12-SEP-1997  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/026,062  
 FILING DATE: 13-SEP-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Verny, Hana  
 REGISTRATION NUMBER: 30,518  
 REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-324-1677  
 TELEFAX: 650-324-1678  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1837 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-928-361B-5

[illegible]

```

QY 412 -----TTPKETA-PTTP-----KKL 425
Db 842 AGQADTSLNLFVQTHKSGLEIDPMVGLPEDPKSGNLVHPYTNQTMGSLSYLAANKL 901
QY 426 T-----PTPEKLAPTTPE-----439
Db 902 TVDTEYGLPDLTLTGTPVLDVSLIPNPETGELFDPSIDELNNGTACIVSGISASES 961
QY 440 ----KPAETPE-----LAPTT-----PEEPTPTTP--EEBAPTTPKAA 474
Db 962 LLSQASAPIDPATNMVGEFGGLNIPATGVMIPLGSLGSEQTSPSELEDGIIPEEVA 1021
QY 475 PPTPEAPPTTPKEAPPTTPE-----PA 498
Db 1022 ANADFKLSIPPSYEPESIPEDOKIDSISELMYDIESGRLIGQYKRPISGIADLNP 1061
QY 499 PTPKETAPTTPKGTAPTTLKEAPATTPPKPAKELAPTTTKEPTSTTSKPAATTTKGT 558
Db 1082 MKTPQTOSVTKRIDPTT---GLPND--PTGHLINFTNNMTDSSFAKAYKAVANSNGI 1136
QY 559 APPT---PKEPAPTTPKEAPATTPKGTAPTTLKEAPATTP-----KKPAK-- 601
Db 1137 KTDVYGLPVDLITGLPDPVSDIPFNSTGELVDPSTGKINNYTAGIYSGKGLPIE 1196
QY 602 ----ELAPTTKGP-----TST-----SDKAPPTPKETAPTTPKEPA--- 636
Db 1197 DENGLDPSKTLPIDGNOLNPEINSTVSGSTGSKRPAGIPVGGGVDPDEAKDQ 1256
QY 637 -----PTTPKBPAPTTPETPTTSEVSTPTTKE--PTTIKS--PDESPEL 681
Db 1257 ADKGDGLIAPPTNSINKDPVTNQTGNTGNIINPETGKVIPLSLGSLNPSFNPDQ 1316
QY 682 SAETTPKALE---NSPKPEGVPTTKTPAATKPEKTTTAKOKTTERDLRTTPTTAAPKM 738
Db 1317 TDEITGKRVDTVGLPYDPSTGELIDPATKLPIDGSAVAGEILTEVINTTDEVTLGP-I 1375
QY 739 TKETATTTETKTESKITATTOVSTTQDTPPKITLTTTLAPVTTTKTITITE- 797
Db 1376 DLEGLRDP-----VSGLRPLNGTLVD-----PANKPTIPGSHS 1411
QY 798 --LNKKEETAKKPRDRTATNSKATTPKPKP--TKAP--KKPTSKKPKTAPRVKPKT--- 849
Db 1412 GFINGTSGEOSHEDKPDSTG---KPLDNTGGLPDEDSGSLINPETGKLGDSHSGT 1465
QY 850 TTPRKATSTMPLELNPTSRISAEAMLOTTTPRNOTPNSKLYEVNPK--SEDA----- 898
Db 1466 MPVPKAGQGENGIMPEQILEAL-----NKLPTSNENVISPRSSDAVDPRTNTWM 1518
QY 899 -----GGAEGETPHML-----RPHVPEVTTPMDVLPVRVNGGIII 936
Db 1519 NKISGOTFOVDGKKTIPGSAASYIHATLGPRTQDPTTGLP--SDPSTGLPFIPEGNVLV 1576
QY 937 NPHLSD-----ETNIC-----NGKPVNGLTTL--RNGTLV-AF 966
Db 1577 DPOTGQIKGSVYSLVYKKNIVTEAAGLVDPDKTGFIDPISYLPFAKNGELIDPI 1636
QY 967 RGHYF 971
Db 1637 SGKYF 1641

```

```

; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 25
; LENGTH: 744
5202236-25

Query Match          9.4%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 2.2e-25;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 51 PKITTAKPINRPSLSPNSDTSKETSILVNNKETTVETKETTTNNKQSTDGKEKTSAKE 110
Db 24 PKMTYPTPKKPSYPP-----TKSKPTV-----KKIT----- 53
QY 111 TQSIKTSAMDAPTSKVLAKPTPKAETTTKGALTTTPEEPTPTTKE---PASTTPE 166
Db 54 -----YPTTK--AKPS-----YPTTKPKTYPTTKPKLTVYPTTKPKP 92
QY 167 PTPPTTKSAPT--TPKEAPATTPKSAPTTPKEAPATTPTKBPAPTTPKEAPATTPKBP 223
Db 93 SYPTTKSKPTTKPKLITVPTTKAKPSYPTTKPKTKTTPPT--YKRLITVPTTKKPSY 151
QY 224 TTTKSAPTTPKEAPATTPKBPAPT--PKBPAPT---PKBPPTTPKBPAPTTPKEP--- 275
Db 152 PTYKPKPSY--PSYTKTKTTPPTTKPKLITVPTTKPKPSYPTTKPKTKTTPPTTKPKL 209
QY 276 -APTTPKEPA--PTAPKKA--PTTPKEPA--PTTPKEPA--PTTPKES--PTTPKKA 324
Db 210 YPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 269
QY 325 -PTTTSAPTTPKEAPATTPKSAPTTPKPSPTTKKEPA--PTTPKEPA--PTTPKPKP 377
Db 270 YPTTKAKPSY--YPTTKAKPSY--PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 322
QY 378 A--PTTPKEPA--PTTPKKA--PTTTPKKA--PTAPKBPAPTTPKEAPATTPKLPPT 429
Db 323 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 382
QY 430 PEKLAETPE--KPAETPEELAPTTPEEPT--PTTPEEPA--PTTPKA--AAPNT-----P 478
Db 383 TYKAKPSYPTTKAKPSY--PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 438
QY 479 KBPAPTTPKEAPATTP---PKBPAPTTPKEAPATTP--PKGAPATTLKBPAPTTPKBPAPK 532
Db 439 SYPTTKPKISYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 498
QY 533 ELAPTTTKEPTSTSDKBPAPTTPKGTAPTTPKEAPATTPKBPAPTTPKGTAPTTLKEPA- 591
Db 499 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 546
QY 592 -PTTKKAPKELAPTTTKGPT--STSDKBPAPTTPKGTAPTTPKEAPATTPKBPAP--PT 646
Db 547 PPTTKAKPSYK--AKPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 601
QY 647 --TPPTPTTSEVSTPTTKE-----PTTIHSPD-----ESTPELSAEP--PKALENS 693
Db 602 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 661
QY 694 KPEGVPTTKTPAA-----TKPEMTTAKDKTTERDRTTPEPTTAP-----KMT 739
Db 662 PSYK--PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 719
QY 740 KETATTTKTESKITATTT 759
Db 720 YKAKPTYPSTYKAKPTYPST 739

```

RESULT 4  
US-09-103-429A-3

```

Sequence 3, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: CDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinisti & Michaels, P.C.
STREET: 118 No. 6187558th 110ga
CITY: Itasca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-3

```

```

Query Match          9.2%; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.1%; Pred. No. 7.6e-25;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

```

```

OY 132 PPKAETTTKGAALTPPEPTTPKPEASTPKPEPTTTIKSAPTTPKPEAPTTTSSAP 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 93 PEPPEAE-TTQAAPAT----QAFTTQAFTTT------QAFTTTQ--ATTTTQAP 135
OY 192 TTPKPEAPTTTKEPAATTPKPEAPTTTKEPAATTTKSAPTTPKPEAPTTPKKAPPTPK 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 TTTO--APTTTQ--ATT----QAFTTTQ--APTTTQAFTTTQ--APTTOQ--APTTOQ- 183
OY 252 PAPTTPKEPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPPAKPKAPATTKEAPATT-PEEPA- 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 -APTTOAFTTTTQAATPAATPAATPAATPAATPAATPAATPAATPAATPAATPAATPAAT 241
OY 309 -----PTTTKEPSPPTPK- 321
DB 242 ELLPNCPCPADFDIHLIPDKYCNLFYQCSNGTYEQRCPEGLYNNPYQRCDSFANVEC 301
OY 322 ---EPAPTTT----- 328
DB 302 DGEISAPRVTEGNEDEDIDIGLLDNGCPANFEIDMLPHGNRCQDKYYQCVHNLVERR 361
OY 329 -----KSAPTTPKPEAPTTTKSAPTTPKPEPSTTTKEP 361
DB 362 CGAGTHSFELQOCHIELVGTCLPGGESEEVADDEA-CTGMWYCTPEITIEWEPLNCGP 420

```

```

OY 362 APTT-----PKP-----APTTKKKAPPT----- 381
DB 421 ADFPSIDHLLPHESDGGYLOQVHGOTIARPCPGNLIHSPATOSCESPYTAGCQVFECDS 480
OY 382 ---PREPAPTTPEKPEAPTTTTPKPAATPEAPTTTPEKTAATPTTKLPTTPEKLAPTT 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 NCGTSTAPTAAPTAAPTAAPTAAPTA--APSTVVPATPATAPVAPPTT---AAPT-- 533
OY 439 EKPAPTTPEELAPTTPEEPTTPTEPEAPTTPKKAAPT---PKPEAPTTPEAPTTPEKE 496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 --PAPTAAPTAAPTAAPESPTTVP--PTAAPTAPTAAPTAAPTEITVY---SAPTAPT 587
OY 497 PAPTTPKEAPTTPKGTAPTTTKEPAPTTPKKAPKEAPTTTKEPTSTSDKAPATTPK 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 588 AAPTAAPTAAPTAAPTAAPTEITVTSPTTAPATTAAP---APNTT-----VTVP 632
OY 557 GAAPTTPKEAPTTPEKAPTTPKGTAPTTTKEAPTTTKEAPTTTKEAPKEAPLAPTTKEPTST 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 633 TAAPT--AAPAPNTTVVPTAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 677
OY 617 DKPAPTTPKETAPTTPKPEAPTTPKK 643
DB 678 ---APATPEDDIDP--PLPNDPINP 699

```

```

RESULT 5
US-08-700-651-5
Sequence 5, Application US/08700651B
Patent No. 6015862
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF CRYPTOSPORIDIUM PARVUM
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
EARLIER FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 5
LENGTH: 1721
TYPE: PRT
ORGANISM: Cryptosporidium parvum
US-08-700-651-5

```

```

Query Match          9.0%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 7.3e-24;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

```

```

OY 5 KKNFTKKKPPKPVVDEAGSGLDNGDFKVTTPDSTGKNNKSTSPKITTAKPINRPS 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 RSNFTK---TTPEPANTYAGVRSN-ETKTEPSANT---NFLVDPK1-----N 158
OY 65 LPPNSDSKSE-----TSLVANKETVTEKETTNNKSTGKEKETSATK 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 159 APCNSSENSFEGQGLFDWGSKVYIPTKCVGKHHTTTTTTTTTTTTTTTTTTTT 208
OY 110 ETOSIEKTSANDLAPTSKVLAKPTPKAETTTKGGALTTPKEPTTPKPEASTTPKEPTT 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
OY 170 TTITSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKEPAATTPKPEAPTTTKEPAATTTTSA 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
OY 230 PTTKPEAPTTPKKAPATTPKPEAPTTTKEPTTPKPEAPTTTKEPAATTPKPEAPTTAPK 289

```

[illegible]

```

US-08-928-361B-6
:
: Sequence 6, Application US/08928361B
: Patent No. 6071518
:
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENT
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: NUMBER OF SEQUENCES: 30
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
:
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: VERNY, Hana
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480.76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
:
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1721 amino acids
: TYPE: amino acid
: STRANDEDNESS:
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08-928-361B-6

```

## RESULT 6

```

Query Match          9.0%:  Score 489;  DB 3;  Length 1721;
Best Local Similarity 22.1%:  Pred. No. 7.3e-24;
Matches 284;  Conservative 82;  Mismatches 476;  Indels 442;  Gaps 41;

QY  5 KKNRTKKRPPRPVVDGASGLDNGDFVLTTPDSTQOHKNVSTSPKTTAKKPINRPS 64
Db  116 RSNETK---TTEPSANTYAGVRSN-EKTEPPSANT---NELVDPKI-----N 158
QY  65 LPNNDTSKE-----TSLVNKEVTVETKEITTNKQSTDOKEKETSAK 109
Db  159 APCNSENSEFEGQGLFDMGSKVYIPYTKCEGVGNHTTTTTTTTTTTTTTTTT 208
QY  110 ETQSIEKTSAKDLAPTSVLAKPRPKATETTTKGALTPPKKEPTPTTKEPBASTTPKEPTP 169
Db  209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
QY  170 TTISAPPTTKEPAPATTTKSAPTTPKEBPATTTKEBPATTTKEBPATTTKSA 229
Db  241 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
QY  230 PTTKEPAPTTPKKPAPTTKEBPATTTPKKEPTPTTPKEBPATTTKEBPATTPK 289
Db  301 TTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 350
QY  290 KPAETTPKEBPATTTKEBPATTTKEPSTTPKEBPATTTKSAPTTPKEBPATTTKSAPTT 349
Db  351 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTTTT 406

```

```

QY 350 PKE-----PSPT----- 357
Db 407 TSETSEVIRKPEDEMCLEKNGEAGATVYVIGKDRIGENGMATMIPNDTHVREPK 466
QY 358 -----TKEP-----ATTKE 368
Db 467 VKDVNTISVRCRKGAKLEPPDRSLDFTIPVACHNSCSIIIVGVSGDKIHVSPGSKD 526
QY 369 PAPTTPKKAPPTPK-----PAPTTPKKAPPTPKPAPAPKEPA 410
Db 527 VSLIS-----APIOPSELENEVCDICTAKYGAHSGYQSADPVTITTKPPTTT--TGA 580
QY 411 PTPPKETAPPTPKLPTTPPEKLAPPTPKPAPPTPEELAPPTPEEPPTTPPEEPAPTP 470
Db 581 PGOPPTTTTGSPEKPTTTTATTTT-----TTLNPIITTTTQKPTTTT-----TT- 627
QY 471 KAAPPTPKPAPPTPKPAPPTPKPAPPTPKPAPPTPKPAPPTPKPAPPTPKPAP 530
Db 628 --KVPKRPPIATTTTTLKPIVTTTTTATVTTTTTTPPT-----TTTKRDEMTITTTPL 680
QY 531 PK-----EIAPTTKEPTS-----TSDKPAPTTPKG-TAPTTPKEPAPT 570
Db 681 PDIGDIEITPIEKLMDKTYMTIYVNSGLLSDNDERIPGSOAQIADTSMLEFVQTH 740
QY 571 PKPAPPTPKGAPATTLKE--PAPTTPKKAPKELAPTTTGGPTTSDKPAPTPKET 627
Db 741 KSTGLPIDPVGILPFPKSGNLVHYPTNOMSGLSVLAAXMLVDTDEYV--LPIDT 798
QY 628 APTTPKEPAPTTPKKAPPTPKPAPPTPKPAPPTPKPAPPTPKPAPPTPKPAP 673
Db 799 LTGPIPLDPSVLIIPN--PEIGELFDPISDINMGITAGIVSGISASESLISQKSAIDPA 856
QY 674 -----PDEST-----PELSAETPKPALENSK 695
Db 857 TNMVYGEFGLNPNATGVMIPELGSEQOTQFSEIEDGIIPEVAAANADKFKLSIP- 915
QY 696 EPQPTTKTKTAATKPEMTTAKDKTTER-----DLRTPEPTTAPAKM 738
Db 916 -PSPV-----ESTPE-----KQOKIDISIELMADIESGLIQVSKRPIPSGIADLNP 963
QY 739 TKETATTTKTESKITATP--TOVTSITTTODTTPFKITTLTKTTTLAPKVTYTKKITYTTE 797
Db 964 IMKPTPTDSTGKPIDPTGLPENPTGHLINPTNNNTMDSBAGAKYKAVSNGIKTDN 1023
QY 798 IMNKP--EETAKPKD-----RATNSKATTPKQKPTKAPKKTSGKKKXTMP----- 842
Db 1024 VYGLPVEIGTGLPKDPSDIPFNSTGTGLVDPTGKPIINNSTAGIVSGKGLPIEDENG 1083
QY 843 -----RVKPKPTPTPKKMTS--TMP----- 862
Db 1084 NLFPDSTNLPIDGNNOVLNPEINTSVGSISGTTKPKPGIIPVNGGVVPEBEAKQADKG 1143
QY 863 -----LNPSTIADAMLOTTTRPNQTPNSKILVENPKSEBDAGAGETPHMLRPHFMP 917
Db 1144 KQGLIIVPTNSINDPVTNIQYSMTGTNI--INP--ETGKVIIPGSLPGSLNPSFNTP 1197
QY 918 EYTPD-----MDYLPRVNOGIITNP-----MLSDETNINCKN----- 950
Db 1198 QOTDEITGKPDVIVGLYDPSTGEIIDPATKLPISGVAGDELITVELNITTDDEVGLP 1257
QY 951 -----PVDGLTTLRNGTLV 964
Db 1258 IDLETGLRDPVSGLPQLPNGTLV 1281

```

```

RESULT 7
US-07-638-431-2
: Sequence 2, Application US/07638431
: Patent No. 5198535
: GENERAL INFORMATION:
: APPLICANT: Hoffman, Stephen L.
: APPLICANT: Charoenvit, Yupin

```

```

APPLICANT: Hedstrom, Richard
APPLICANT: Khumitch, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: Immunogen and gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: A. David Spevack
STREET: NMRC Building 1 T-12 National Naval
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/638,431
FILING DATE: 19910110
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avrom D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-638-431-2

Query Match 9.0%; Score 488.5; DB 1; Length 826;
Best Local Similarity 26.4%; Pred. No. 3.5e-24;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

QY 168 TPTTISAPTTTKEPA-PTTTSAPTTTKEPA-PTTTEP-----APTTPKEP-APTTPKE 220
Db 273 TPKKVDCCOILPIPIVPIPKIKIEKPSNPEEYVNDPNDPNPNPNPNPNPNPNPN 332
QY 221 PAPTTPKSAPTTPKEP-APTTPKKP-APTTPKEP-APTTPKEP-TPPTTKEPAPTTPKEPA 276
Db 333 PNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 392
QY 277 PTTTPKEPAPAPAK--KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 334
Db 393 RNNPKRRNPKPKPKPNPKPNPKPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 443
QY 335 TKEPAPTTPKSAPTTPKEPSPTTKEPAPTTPKEPA-PTTTPKKP-APTTPKEPAPTTPKE 392
Db 444 PKNPNP-----NEPSNNKPNPN-----EPLNPNPEPSNPNAPSNPNE--PSNPNE 492
QY 393 PAPTTPKKPAPTAPKEP-APTTPKETAPTTPKKLITTTTBEKLAPTTPPEKAPPTPEELAP 451
Db 493 PS-----NPNPEPSNPNEPSNPNE--PSNPK-----PSNPNE--P 523
QY 452 TTPPEEPTTPPEEPAPTTPKAAAPNTPKEP-APTTPKEPAPTTPKAPAPTTPKETAAPT 510
Db 524 SNPNPE--PLNPNP-----SNPNPEPSNPNEPSNPNE--PSNPKP--PSNPNE----- 564
QY 511 KGTAPTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSDKAPPTTGTAPPTTPKEAPPT 570
Db 565 -----PSNPPEPNPEE--PSNPKEP-----SNPEEPINPELNPKEPSNPEESN 606
QY 571 PKPAPPTPKGAPATTLKEPAPTTPKPKAPKELAPTTTGGPTTSDKAPAPTTPKETAPT 630
Db 607 PKPPI-----NPEESNPKP-----INPEDNENPLIIODEPIEPRNDSNVIP 649

```



```

QY 631 TPKEP-----APTTKKPAPPTTPEPTTSEVSTPTTKEPTTIHKSPESTPELSAE 684
DB 650 LPIIFQKGNMIPSNLPENPSDSSEVEYPRPNDNGENSNNTMKSKNKNI---PNEPIPSGDN 706
QY 685 PTKALENSPK-----EPGVPTTK 703
DB 707 PYKGHEERIPKPHRSNDYVYDNNVKNKNKDEPEIPNNE 745

RESULT 8
PCT-US92-00018-2
: Sequence 2, Application PC/TUS9200018
: GENERAL INFORMATION:
: APPLICANT: Hoffman, Stephen L.
: APPLICANT: Charoenvit, Yupin
: APPLICANT: Hedstrom, Richard
: APPLICANT: Khushf, Srisai
: APPLICANT: Rogers IV, William O.
: TITLE OF INVENTION: Protective malaria sporozoite surface protein
: TITLE OF INVENTION: Immunogen and gene encoding
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: A. David Spevack
: STREET: NMDC Building 1 T-12 National Naval
: STREET: Medical Center
: CITY: Bethesda
: STATE: MD
: COUNTRY: USA
: ZIP: 20814-5044
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PC-DOS/MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/00018
: FILING DATE: 19920103
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Spevack, Avram D.
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 295-6759
: TELEFAX: (301) 295-4033
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 826 amino acids
: TYPE: AMINO ACIDS
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US92-00018-2

Query Match 9.0%; Score 488.5; DB 5; Length 826;
Best Local Similarity 26.4%; Pred. No. 3.5e-24;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

QY 168 TPTTISAPTTTPEKA-PTTTSAPTTTPEKA-PTTTEK-----APTTPKEP-APTTPKE 220
DB 273 TPOKVDQCOIPIPIYIPKIKPEKPSNPEEPVNPDPDNPNNPNNNNNNNNNNPNN 332
QY 221 PAPTTSKAPTTPEK-APTTPKEP-APTTPKEP-APTTPKEP-APTTPKEPAPTTPKEA 276
DB 333 PNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 392
QY 277 PTTTKEPAPTAPK--KPATTTKEPAPTTPKAPATTTTKEPAPTTPKEAPTTPKSAPTT 334
DB 393 RNNPKRNNPKKPKNNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPN 443
QY 335 TTEPAPTTPKSAPTTPKEPAPTTPKEAAPTTPKEA-PTTPKEP-APTTPKEPAPTTPKE 392
DB 444 PAKPMP-----NPSNPNKPNPN-----EPNPNPNPNPNPNPNPNPNPNPNPNPN 492
QY 393 PAPTTPKAPAPAPKAP-APTTPKEAPTTPKAPPTTPKEKLAAPTTPKEPAPTTPKEELAP 451

```

```

DB 493 PS-----NPNESNPNESNPNNE--PSNPKK-----PSNPNNE--P 523
QY 452 TTEEPPTTPTEEPAPTTPKAAPNTPKEP-APTTPKEPAPTTPKEPAPTTPKEAPTTP 510
DB 524 SNPN--PLNPNP-----SNPNESNPNESNPNNE--PSNPKK--PSNPNNE----- 564
QY 511 KGTAPTTLKEPAPTTPKPAKELAPTTPKEPTSTSDRAPPTTGTAPPTTPKEAPTTP 570
DB 565 -----PSNPEEPNPEE--PSNKEP-----SNPEEPINPELNPKEPSNPEESN 606
QY 571 PKEPAPTTPKGTAPTLKEPAPTTPKPAKELAPTTPGTSTSDKAPPTTPKETAAPT 630
DB 607 PKEPT-----NPNESNPKP-----INPDNPNPLIODEPIEPRDSDNVIPI 649
QY 631 TPKEP-----APTTKKPAPPTTPEPTTSEVSTPTTKEPTTIHKSPESTPELSAE 684
DB 650 LPIIFQKGNMIPSNLPENPSDSSEVEYPRPNDNGENSNNTMKSKNKNI---PNEPIPSGDN 706
QY 685 PTKALENSPK-----EPGVPTTK 703
DB 707 PYKGHEERIPKPHRSNDYVYDNNVKNKNKDEPEIPNNE 745

RESULT 9
5202236-13
: Patent No. 5202236
: APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
: SUSAN L.; MCCANDLISH, RUSSELL; TENA; FILIPULA, DAVID
: TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
: PROTEIN
: NUMBER OF SEQUENCES: 39
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/528,762
: FILING DATE: 25-MAY-1990
: APPLICATION NUMBER: 82,456
: FILING DATE: 07-AUG-1987
: APPLICATION NUMBER: 933,945
: FILING DATE: 24-NOV-1986
: APPLICATION NUMBER: 650,128
: FILING DATE: 13-SEP-1984
: SEQ ID NO:13:
: LENGTH: 652
: 5202236-13

Query Match 8.8%; Score 476.5; DB 6; Length 652;
Best Local Similarity 29.7%; Pred. No. 1.6e-23;
Matches 221; Conservative 71; Mismatches 284; Indels 169; Gaps 45;

QY 51 PRTTAKPINRPSLPNDSOTSKETSLVNNKETVETTTTNNKOSTDKEKTTSAKE 110
DB 11 PKMTPTPTTKKPSYPP-----TYKSPY-----KPKIT----- 40
QY 111 TOSIEKTSKADLAPSKYAKPTPKAETTTGAPALTTPKEPTTPPK-----PASTTPKE 166
DB 41 -----YPTTYK--AKPS-----YPTTYKPKTYPTTYKPKLTTPPTYPKPK 79
QY 167 PTPPTKSAPT-TKEPAPTTPKSAPTTPKAPATTTTKEPAPTTPKAPATTPTPKEA--P 223
DB 80 SYPTPKSKPTTKKLTTPPTYPKAKPSYPTTPKPKKTYPPPT-YKPKLTTPPTYPKAKSYP 138
QY 224 TTTTSAPTTPEKAPATTPKAPAT-TKEPAPT-----PKPPTTPTEKAPATTPKEAPT 278
DB 139 PTKPKPSYP--PSYKTKKTYPTTPKPKLTTPPTYPKPKPSYKPKKTYPTTPKPKLT 196
QY 279 TPKEPAPTAPKKA--PTTPKEA--PTTPKEAPTTPKEPSPTP-----KEPAPTTPK 329
DB 197 YP-----PTYKAKPSYPTTPKAKPSYPTTPKAKPSYPTTPKAKPSYPLIKAKPSYPT 252
QY 330 SAPTTTKEP-----PTTTSAPTTTPKEPSPTTTPKEA--PTTPKEA-----PTTPK 376
DB 253 AKPTPKKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 308

```

Qy	377	PA-PTTPKEBA--PTTPKEBAPTTPKKAPAKPEAPTPKEAPTPPKLPTTPPK	432
Db	309	PSYPTPKAKSPYPTPKAKSPYPTPKAKSPY----	PKAKSPYPTPKAKSPY- YK 362
Qy	433	LAPTPPE--KPAPTPPEELAPTPPEEPT--PTTPPEBA-PTTPKAAPNTPKEBAPTTPK	487
Db	363	AKPYPTSPYKAKPSYP----	PTYKAKSPYPTPKAKSPYPTPK- AKPSYP-----PTYKA 413
Qy	488	EPA-PTT---PKEPAPTTPKEAPTP--PKGAPPTTLKEBAPTTPKKAPKELAPTTPK	540
Db	414	KPYPTSTYKAKSPYPTPKAPISYPTPKAKAKSPYSTYKAKSSYPTPKAKKSPYPTPKA	473
Qy	541	EPTSTSDKAPAPTPKGAAPTTPKEBAPTTPKEBAPTTPKGAPTLKEBA--PTTPKKP	598
Db	474	KPYPTSTYKAKPYTK--AKPSYP----	PTYKAKS-----YPTPKAKKSPYPTPKAKP 521
Qy	599	APKELAPPTTKGPTSTYSDKAPAPTPKEBAPTTPKEBA--PTTPKKPA-PTT-----PETP	651
Db	522	TYK-AKPYTKKAKSPYPTPKAKKPSYP----	PTYKAKSPYPTPKAKKPYPTSTYKAKKPSYP 576
Qy	652	PEPTSEVSTPTTPTEPTTIHKSPEBSIPBELSAETPPALENSPEBOPVPTTKPAATKPE	711
Db	577	PTYKAKKPSYPPT-----YKAKSPYPTPKAKPYTKA----	KPYPTST-----YKAKPS 620
Qy	712	MTTKAKTTERDLTKTPETTTAAP	736
Db	621	YPTPKAKPYTKAKSPYPTPKAKP	645

```

RESULT 10
US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELIE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
; US-09-219-849-5

```

	Query Match	8.3%	Score 452;	DB 4;	Length 960;
	Best Local Similarity	25.6%;	Pred.No. 9.3e-22;		
	Matches 173;	Conservative	60;	MisNo. 134;	Gaps 41.
QY	132 PTPKAEITTKGALATPKKEPTPTTPEKPASTAPKEEPTTIISADPTTKEPAP--TTTKS	189			
	: : : :			: :	
Db	72 PGPGAGPSNDPG--PGCAPGPAGP--PSSRDPGGGAGC-PAGPPGSNDPGPGAPGGA	126			
	: : : :			: :	
QY	190 APTTPKEPAPTPTTKEDPA-TTTPKEPAPTTTKEBAPTITKSAPTTPKEAPTTD--KTGA	245			
	: :   : : : :			: :	
Db	127 GPPGSDNPPGAPGAPGPGSGNDPGPCGAPGAPGPGSGNDPGPCAPAPGAPGPGSGNDPG	186			

QY	246	PTTKEPAPITP---	KEPTITTTKEPA-	PTTKEPAPITTTKEPA-----	284	
Db	187	P--PGAPGPAGP	PGSRRDGP	PGAGP	PGSGRPGGAGHGPAGKGAHGAPGPGGAGH	244
QY	285	-----PAPK-K	PAITP---	KEPAPITTTKEPA-PTTKEPSPTTTKEPAPITTKS	330	
Db	245	PAGKGPAGP	PGPGAGPAGP	PGSRRDGP	PGGAPGAPGPGSGSDGP--PGAGP----	298
QY	331	APTPTKEPAPITTKS	APTTPKEPSPTTTKEPAPITTK	KEPAPITP---	KKPAPITTTKEPAP	387
Db	299	GPSPGRDGP-----	PGAPGPAGP	PGSRRDGP--PGAPGPAGP	PGSGRDGP--PGAPGP	348
QY	388	TTP---	KEPAPITTTKPA-	-----PTAPKKEPAPITTTKEPAPITTPPKLTPTTEKLAPT	437	
Db	349	AGPSPGRDGP	PGGAPGAPGPGSGSDGP	PGGAPGAPGSGSDGP----	PGAGPAGP	404
QY	438	PEKPAPT--TPELA	APTTPPEETPTTPEEPAPITTKAA-	PTTPKEPAPITP---	KEPAPT	492
Db	405	SRDPPGPAGP	AGPSPGRDGP--PGAPGPAGP	PGSGSDGP	PGAPGPAGP	462
QY	493	TPKEPA-----	PTTPKET-A	PTTPKGT-----APTTLKEPAPITTTKKP-APRELA	535	
Db	463	GAGPAGP	KGAGPAGP	KGAGPAGP	KGAGPAGP	522
QY	536	PTTTPKEPTSTTDK	APITTPKGA-PTTPKEPAPITP---	KEPAPITTKGA-PTTLKEP		590
Db	523	PGPSRRDGP	PGGA-PGAPPGSRRDGP	PGGAPGAPGSGSDP	PGGAPGAPGSGSDP	580
QY	591	APTTPKEPAP-----	KELAPITTKGPTST--	TSDKAPITTPKETAPITTTKEPAPITPK	642	
Db	581	GP--PGAPGPAGP	PGSRRDGP	PGGAPGPAGP	PGSGSDGP--PGA	636
QY	643	PAPTTP-----	ETTPPTTSVSTPTTTKETTT	HKSSDSTPELSAPPKALEN-----	S	693
Db	637	PGGAPGP	SGSDGP	PGGAPGP	SGSDGP	699
QY	694	PKPGVPTTKTPATK	709			
Db	696	SRDPPGPAGP	AGPAGK	711		

RESULT 11  
US-08-479-537A-5  
Sequence 5, Application US/08479537A  
Patent No. 5861381  
GENERAL INFORMATION:  
APPLICANT: CHAMSON, Pierre  
APPLICANT: KIENY, Marie-Paule  
APPLICANT: LATHE, Richard  
APPLICANT: HAREVEUIL, Maira  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TITILE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22113-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,537A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 90/13101  
FILING DATE: 23-OCT-1990  
PRIOR APPLICATION DATA:

```

1 APPLICATION NUMBER: WO PCT/EP91/00835
2 FILING DATE: 23-OCT-1991
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 08/039,320
5 FILING DATE: 04-APR-1993
6 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: US 08/403,576
8 FILING DATE: 14-MAR-1995
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Teskin, Robin L.
11 REGISTRATION NUMBER: 35,030
12 REFERENCE/DOCKET NUMBER: 017753-025
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: (703) 836-6620
15 TELEFAX: (703) 836-2021
16 INFORMATION FOR SEQ ID NO: 5:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 1867 amino acids
19 TYPE: amino acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: peptide
23 FEATURE:
24 NAME/KEY: Peptide
25 LOCATION: 128..1727
26 OTHER INFORMATION: /note="The amino acids spanning
27 OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
28 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
29 OTHER INFORMATION: repeats varies from 1 to 40."
30 FEATURE:
31 NAME/KEY: Peptide
32 LOCATION: 134
33 OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa
34 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
35 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
36 FEATURE:
37 NAME/KEY: Peptide
38 LOCATION: 144
39 OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa
40 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
41 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
42 FEATURE:
43 NAME/KEY: Peptide
44 LOCATION: 147
45 OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa
46 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
47 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
48 FEATURE:
49 NAME/KEY: Peptide
50 LOCATION: 1..21
51 OTHER INFORMATION: /note="Amino acids 1 to 21 are a
52 OTHER INFORMATION: 21 amino acid precursor sequence."
53 US-08-479-537A-5

```

Query Match	7.8%	Score 424.5	DB 2	Length 1867
Best Local Similarity	23.7%	Pred No. 1.1e-19		
Matches 225	Conservative 69	Mismatches 511	Indels 241	Gaps 53
QY	75	TSLTVAKETTVKEKTTTNTKQSTDQGEKTSKAKETQSIEKTSKAD-LAPTSKVLAKPT	133	
Db	16	TVLTV-----VTSQHAASSTPGGEKETSATQSSVSPSSKNAKMSMTSSVSSAS	65	
QY	134	P-KAETTTGCP--ALTPKRP-----PPTTKEPASTTTPKPEPTTTKISAP	176	
Db	66	PGSGSSTTQGGODVTLAPATERPAGSAATGODVTSVPTLRPAGSTTPRAHDVT---SAP	122	
QY	177	TTTPKEPAPTTT-----KSAPTTPKPE-----APTTPKEAPTPPKPEAPPT	217	
Db	123	--DNKPAPESTAPXAHGVTSA PDKRPXGSGSTAPXAHGVTSA PDKRPXGSGSTAPXAHGVT	180	
QY	218	TKE--PAPTT-----KSAPTTPKPEAPTPK-----KAPTTKEPAPTPPKPEPTPT	264	

[illegible]

RESULT 12  
US-09-083-116-5  
Sequence 5, Application US/09083116  
Patient No. 6203795  
GENERAL INFORMATION:  
APPLICANT: CHAMBER, Pierre  
APPLICANT: KIENY, Marie-Paule  
APPLICANT: LATHE, Richard  
APPLICANT: HAREUVEN, Mara  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DONNE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22131-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,116  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/479,537  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA: US 08/039,320  
APPLICATION NUMBER: 04-APR-1993  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1867 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1727  
OTHER INFORMATION: /note= "The amino acids spanning  
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat  
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
OTHER INFORMATION: repeats varies from 1 to 40."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA  
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
OTHER INFORMATION: 21 amino acid precursor sequence."  
US-09-083-116-5

Query Match		7.88;	Score 424.5;	DB 4;	Length 1867;
Best Local Similarity		23.78;	Pred. No. 1.1e-19;		
Matches 255;		Conservative 69;	Mismatches 511;	Indels 241;	Gaps 53;
Qy	75	TSLVNKKETVETKETTNNKOTSDGKETTSAKETQSIKETSAND-LAPISKVLAKPT	133		
Db	16	TVLIV-----VYSGSHASTPGGEKETSATQSSVSSSTERNAAVSMSTSVLSHS	65		
Qy	134	P-KAETTTKGP--ALITPKPE-----TTTKEKESATPKPEPTTTIKSAP	176		
Db	66	FGSGSTTQGGDVLAPATPEPASGAATWQDVTSVETPRALGSTITPRAHDT-----SAP	122		
Qy	177	TTPKPEAPTT-----KSAPTTPKPE-----APTKEKAPTTKEKAPTT	217		
Db	123	--DNKPAPGSTAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	180		
Qy	218	TKE--PAPTT-----KSAPTTKEKAPTTK-----KRAPTTKEKAPTTKEPTPT	264		
Db	181	APDXRPXGSTAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	240		
Qy	265	PKE--PAPTKEP-----APTKEKAPTTAPK-----KRAPTTKEKAPTTKEPAPT	311		
Db	241	APDXRPXGSTAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	300		
Qy	312	TKEPSTTTKEKAPTT--TKSAPTTKEKAPTT-----TKSAPTTKEKAPTTTK-----E	360		
Db	301	APDXRP--XPGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	359		
Qy	361	PAPTTKEKAPTTK-----KRAPTTKEKAPTTK-----EPAPTTKKRAPTAKEKAPT	412		
Db	360	SAPDXRPXGSTAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	419		
Qy	413	TPKETAP-----TTPKLPTTPEKLAPTTPEKAPTTPEELAPTPEPTPTTPEEAPT	468		
Db	420	SAPDXRPXGSTAPXAHGVT-----APDXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	471		
Qy	469	TPKA-----AAPNTKEKAPTTKEKAPTTKEKAPTTKEKAPTTKEKAPTTKEKAP	523		
Db	472	APXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTA	530		
Qy	524	TP-----KKPAPELAPTTTKEPTSTSDKPAE--TTPKGTAPTTKEKAPTT	571		
Db	531	TAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTA	585		
Qy	572	KEKAPTTKGS-----TAPTTKEKAPTT-----KKPAPELAPTTTGPSTTSD	617		
Db	586	PXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTA	644		
Qy	618	KPAE--TPK-----ETAPTTKEKAPTT-----KRAPTTPEPTTSEVSTP	661		
Db	645	RPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTA	702		
Qy	662	TTTKEPT-----IKSPDESTPELSAPTKEALENSKEGVTTTTPAA-----	707		
Db	703	DXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTA	762		
Qy	708	--TKPEMTTKDK-----TTERDLRTTPTTAAFKMTKETATTEKTESKITATTTQVTS	763		
Db	763	DXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTA	820		
Qy	764	TTTODTTTPEKLTTLTKTTLAPKVTTKKTTTTEIIMKPEETAKPKDRATNSKATTPKQ	823		
Db	821	APDXRPXG-----STAPXAHGVT-----SAPDXRPXGTAAPXAHGVTSAPODXRPXGTA	868		
Qy	824	KPTAPKPTSTTKPKTTPRKRKKTTPPKMTSTPELNPSTRIAEAMLQTTTRNQT	883		
Db	869	GSTAPXAHGVTSAPODXRP--XPGTAAPXAHGVTSA--PODXRP-----PGST	911		
Qy	884	--PNSKLVENPKSEDAEGEETPHMLLRPHVEMEVTPDMQDYLPRVNOGIIINPLMSD	942		
Db	912	APXAHGVTSAPODXRPXGTAAPXAH-----GVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTA	957		
Qy	943	ETINIGKRPVDCGLT--LKNGLVAFRGHYFWMLSPFSPSPARITETVWGIPIPID	997		

Db 958 VTSAPDXRPGSTAPXAHGVTSA-----PDXRPGSTAPXAHGVTSA 1003

RESULT 13

US-08-479-537A-2

Sequence 2, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMBER, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREVENT, Marc

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teakin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1899

OTHER INFORMATION: /note= "The amino acids spanning

OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat

OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such

OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "amino acid 134 is X1 - Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note= "amino acid 144 is Y - Xaa

OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note= "amino acid 147 is X2 - Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note= "amino acids 1 to 21 are a

OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 7.8%; Score 424.5; DB 2; Length 2035;

Best Local Similarity 23.7%; Pred. No. 1.2e-19;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

Db 75 TSLTVNKETTVETKETTNTKOTSDGKKEKTSAKETOSIEKTSAMD-LAPTSKYLAKPT 133

16 TVLTV-----VTGSGHASSTPGKEKETSATDORSSPSTEKNAMSVTSLSHS 65

QY 134 P-KAETTTGP--ALTTPKP-----TPTPKKEPASTTPKEPTTIKSAP 176

Db 66 PGSGSSTGOGOVTLAPATEPASGSNAITGQDVTSVPTVRPALGSTTPPAHVT---SAP 122

QY 177 TTPKEPAPTTT-----KSAPTTKEP-----APTTTKEPAPTTTPKEPAPTT 217

Db 123 --DNKRAPGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 180

QY 218 TKE--PAPTT-----KSAPTTKEPAPTTPK-----KPAPTTKEPAPTTPKPEPTT 264

Db 181 APDXRXPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 240

QY 265 PKE--PAPTTKEP-----APTTKEPAPTPAK-----KPAPTTKEPAPTTTPKEPAPTT 311

Db 241 APDXRXPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 300

QY 312 TKEPPTTPKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTPKPEPTT-----E 360

Db 301 APDXRP--XGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 359

QY 361 PAPTTTPKEPAPTTPK-----KPAPTTTPKEPAPTTPK-----EPAPTTTTPKAPAPTPKEPAPTT 412

Db 360 SAPDXRXPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 419

QY 413 TPKEPAPTTTTPKKTLPPTPEKLAAPTTTPKEPAPTTPEELAPTTPEEPPTTPEEPAPT 468

Db 420 SAPDXRXPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 471

QY 469 TPKA-----AANTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 523

Db 472 APXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 530

QY 524 TTP-----KKPAKELAPTTTKEPTSTSDKRAP--TTPKGAAPTTTPKEPAPTTTP 571

Db 531 TAPXAHGVTSA PDXRPXGSTAP--XAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 585

QY 572 KEPAAPTTTPKGS-----TAPTTTKEPAPTT-----KKAPPELAPTTTGGTSTSD 617

Db 586 PXPGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 644

QY 618 KPAP--TTPK-----ETAPTTTPKEPAPTT-----KKAPPTPEPTTPEPTTSEVSTP 661

Db 645 RXPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVT 702

QY 662 TTPKEPTT-----IKSPDESTPELSAETTPALENSPKREGVPTTKTPAA-----707

Db 703 DXRXPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA PDXRPXGSTAPXAHGVTSA 762



```

Db 472 APXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSAPDXRXPXGS 530
QY 524 TTP-----KKPAKELAPTTTKPTSTSDKRP--TTRGTPPTPKRAPPT 571
Db 531 TAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPXGS 585
QY 572 KEPAPTTPKPG---TAPTTLKEPAPTTP-----KKPAKELAPTTTKGPTSTSD 617
Db 586 PXPGSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDX 644
QY 618 KRAP--TTPK-----ETAPTTKEPAPTTP-----KKPAPTTPPTTSEVSTP 661
Db 645 RXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPXGS--TAPXAHGVTSAP 702
QY 662 TTKERTT-----IHKSPESTPELSAERTPKALENSKPEPEVPTTKTAA----- 707
Db 703 DXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAP 762
QY 708 -TKPEMTTAKDK---TTERDLRTTPETTTAARKMTKETATTEKTESKITATTTQVTS 763
Db 763 DXRXPSTAPXAHGVTSAPDXRXPST--APXAHGVTSAPDXRXPSTAPXAHGVTS 820
QY 764 TTTODTPEKTTTLTKTTLAKVTTTKTTTTEIMMKPEETAAPKDRATNSKATTPKQ 823
Db 821 APDXRXPXG-----STAPXAHGV-----SAPDXRXPSTAPXAHGVTSAPDXRXP- 868
QY 824 KPTAKPKPTSTKKRPMRVRKPTTPPKMTSTMPELNPTSRILAEMLQTTTRNOT 883
Db 869 -GSTAPXAHGVTSAPDXR--XPGSTAPXAHGVTSAPDXRXP-----PGST 911
QY 884 -PNSKLEVPNPKSDAGAGEETPHMLLRPHVEMVEVTPMDVLRPNVNOGIIINPLSD 942
Db 912 APXAHGVTSAPDXRXPSTAPXAH-----GVTSAPDXRXPXGSTA---PXAHS 957
QY 943 ETNINCKRPVGLTT--LRNGTLVAFRGHYFMMLSPFSPAPARITTWKIIIPSID 997
Db 958 VTSAPDXRXPXGSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPD 1003

```

```

RESULT 15
US-08-276-967-2
; Sequence 2, Application US/08276967
; Patent No. 5851817
; GENERAL INFORMATION:
; APPLICANT: Hardy, Daniel M.
; TITLE OF INVENTION: Species-Specific Egg-Blinding Proteins of
; TITLE OF INVENTION: Sperm
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,967
; FILING DATE: Submitted Herewith
; CLASSIFICATION: 350
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: US/SD:418\KIT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679

```

```

; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-276-967-2

```

```

Query Match 7.7%; Score 419.5; DB 2; Length 2476;
Best Local Similarity 32.5%; Pred. No. 3.2e-19;
Matches 163; Conservat 43; Mismatches 177; Indels 119; Gaps 31;

```

```

QY 406 KPEAPPTPKETAPPTPKKLTPTTPEKLAETPEKRAPPTPEELAPTTPEEPPTPEEP 465
Db 313 PSETSVSTKRPVAPTE-----KPTVPSKITITPEKPMVHMEKPIVHT--EKPTVPT-EKP 365
QY 466 APTTPKRAAANTPEKA---PTTPKEAPPTPKETAPPTPKETAPPTPKETAPPTLKP 521
Db 366 TITPEKSTVPT--KKPTVEKEPTLPPE-GPTVPAE-RPTTPREGPAVPPKG--PTVLTE- 418
QY 522 APTTPKKPAKELAPTTTPEKSTSDKRAPPTPKGTAPT--TPKEAPPTPKETAPPTPK 580
Db 419 -----WPTSHTEKSTVHTKPLPLPGKSTIPTEKPMVPTKRT-----TTP- 458
QY 581 GTAPTTLKEAPPTTPKKPAKELAPT--TKGPTST---SDKRAPPTPKETAPPTPEKA 636
Db 459 -TERTIIPAKKPIV---PIEKPMVPTERTIPTERTIPTTEKPTVTEKLIPT--EKPI 512
QY 637 PTTPKKRAPPTTPPTPP-----TISEVSTPT-----TKEPTTIHKSPESTPELSAEP 686
Db 513 VPIEKPIVPEKHTIPTTEKLVLTERTTPTERTTIPTTEKPMVPEKPSVPT-EKPTVPT 571
QY 687 PKALENSKPEPEGVPTTKTPAATKPEMTTAKDTEERDLRTTPETTTAARKMTKETATT 746
Db 572 -----DEPTIPEKLIPT--ERTTTPKRTITPTITRTPTI-----RTTPT 613
QY 747 EKTTEKITATTTQVYSTTODTTPPKITTLTTLAPRVTTTKTTTTEIMNKPEETA 806
Db 614 ERTT-----TPIRTTPTTERTT--IPTKKTT-----VPEKTIIPT-----ERTI 652
QY 807 KPKDRATNSKATTPPKPKOKPTKAPKPKTKKPKTMPRVAKKPTTPPKMTSTMPELNPT 866
Db 653 AP-----TTPQF-SPTLVPTQPAVVPST-----SATVTPRTTIIASCP---PN 693
QY 867 SRIAEMLQTTTRNOTPNSKL 888
Db 694 AHFERCACPVSQ-SPTPNCCEL 714

```

```

Search completed: April 26, 2002, 16:17:34
Job time: 245 sec

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:18:59 ; Search time 82.85 seconds  
(without alignments)  
978.271 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1263

Perfect score: 5698  
Sequence: 1 VKDNKNRTKKRTPKPPVY.....VAALSTAKYKNWPESVFFK 1064

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR:68:\*  
2: PIR:2:\*  
3: PIR:3:\*  
4: PIR:4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062.5	18.6	3020	2	A49332
2	950	16.7	1664	2	T18262
3	853	15.0	1489	2	T31108
4	802	14.1	1274	2	T16251
5	789.5	13.9	1367	1	S48478
6	784	13.8	1188	2	S49915
7	776	13.6	2187	2	T30826
8	672	11.8	1344	1	A35175
9	671	11.8	1151	2	T18535
10	659.5	11.6	1229	2	T25697
11	633	11.1	489	2	T11622
12	632.5	11.1	7962	2	T38346
13	632	11.1	761	2	T34513
14	629	11.0	3507	2	T84672
15	626.5	11.0	990	2	T51618
16	625	11.0	6642	2	T29757
17	622.5	10.9	839	2	T75518
18	622	10.9	3570	2	T45025
19	620.5	10.9	971	2	T19431
20	607.5	10.7	801	2	T29018
21	607.5	10.7	924	2	S79232
22	605	10.6	379	2	S50125
23	559.5	9.8	350	2	S22456
24	556.5	9.8	856	2	T16543
25	556.5	9.8	875	2	S23760
26	555.5	9.7	1630	2	A53577
27	553	9.7	2232	2	T14434
28	551	9.7	620	2	S06733
29	545	9.6	873	2	A47283

30	542	9.5	369	2	S20500	hydroxyproline-ric
31	538.5	9.5	416	2	JU0465	extensin precursor
32	533	9.4	1087	1	QPM5H	neurofilament trip
33	532	9.3	756	2	T27642	hypothetical prote
34	530.5	9.3	1162	2	JH0557	exo-alpha-sialidas
35	530	9.3	865	2	A47282	calcium-binding pr
36	522	9.2	328	2	JU0985	hypothetical prote
37	522	9.2	1459	2	T32271	hypothetical prote
38	518	9.1	813	2	S70795	membrane glycoprot
39	518	9.1	866	2	T45462	neurofilament trip
40	518	9.1	1072	1	A37221	membrane glycoprot
41	512.5	9.0	867	2	T45463	neurofilament trip
42	509.5	8.9	662	2	A45155	membrane glycoprot
43	508	8.9	1832	2	T31113	mucin FIM-C.1 - Af
44	505	8.9	700	2	A54641	mucin-like glycoprot
45	504.5	8.9	606	2	A43427	interpersed repea

#### ALIGNMENTS

RESULT 1  
A49332  
mucin 2 precursor, intestinal - human (fragments)  
N.Alternate names: mucin SMUC-41  
C.Species: Homo sapiens (man)  
C.Date: 10-Mar-1993 #sequence-revision 12-Apr-1996 #text-change 05-Nov-1999  
C.Accession: A49663; A45106; B45106; A43932; B35332; A61257; P00328; P00329  
R.Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A.Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of c  
A.Reference number: A49663; MUID:94132002  
A.Accession: A49663  
A.Molecule type: mRNA  
A.Residues: 1-639 <GU1>  
A.Cross-references: GB:L21998  
R.Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A.Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
A.Reference number: A45106; MUID:93016075  
A.Accession: A45106  
A.Status: not compared with conceptual translation  
A.Molecule type: mRNA  
A.Residues: 626-1895 <GU2>  
A.Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396  
A.Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A.Accession: B45106  
A.Status: not compared with conceptual translation  
A.Molecule type: mRNA  
A.Residues: 2037-3020 <GU3>  
A.Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398  
A.Experimental source: colon  
A.Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R.Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,  
J. Clin. Invest. 88, 1005-1013, 1991  
A.Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym  
A.Reference number: A43932; MUID:91358717  
A.Accession: A43932  
A.Molecule type: DNA  
A.Residues: 1433-1350, 'U', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A.Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864  
A.Note: sequence inconsistent with the nucleotide translation  
A.Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)  
R.Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A.Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl  
A.Reference number: A33532; MUID:89197956  
A.Accession: B33532  
A.Molecule type: mRNA  
A.Residues: 1916-2193 <GU4>  
A.Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874  
A.Experimental source: Intestine  
R.Jenny, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.



```

Db      865  SDEPTPSDEPTPS--DEPTPS--DEPTP-SDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTP 919
QY      302  TTPKEBA--PTTKEESP-TPEKEAPTPTTKS-APTTPKEBAPT--TTKSAPTPEEPS 354
Db      920  SDEPTPSDEPTPEDEPTPEEPIPTDTPSDEPTPSDEPTPSDEPTPEDEPTPSDEPTPSDEPT 979
QY      355  ---PTTKEBAPT--TPKEBAPTTPKKPAATTPKEBAPTTPKKPAATTTTKKAPTAKP 409
Db      980  PSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEP 1039
QY      410  APT---TPKEBAPTTPKKLPTTPKEBAPTTPKEBAPT---TPKEBAPTTPPEEPTP-TP 462
Db      1040  TSPDEPTPE---TPPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSD-EPTPSDEPTSEPT 1094
QY      463  EEPAPTTPKAAAPTTPKEBAPTTPKPA--PTTPKEBAPT-TPKEBAPTTPKGTAPTTPK 519
Db      1095  EEPAPTPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSD 1154
QY      520  EEPAPTTPKPAKELAPTTTKEPT-STSGDKAPAPTTPKGTAPTTPKEBAPTTPKKEBAPT 578
Db      1155  EPTPS--DEPTPSD-EPTPSDEPTSEPTPEEPIPTDTPSDEPTPSDEPTPS--DEPTPS- 1208
QY      579  PKGTAPTTPKEBAPT-TPKKPAKELAPTTTGGPTTSDKAPT--TPKETAPTTPKE 634
Db      1209  ---DEPTPSDEPTPSDEPTPEEPI-----PTDTPSDEPTPSDEPTPSD-EPTPSDE 1253
QY      635  PAPTTPKKAP-TPEPTPEPTSEVSTPTTKEPTTIHKSPESTPELAEPTPKALENS 693
Db      1254  PTPS--DEPTSEPTPEEPIPTDTPSDEPTPSDEPT---PSDEPTP--SDEPTPSDEPTP 1305
QY      694  PREGPVPTTKTPAAT--KPEMTTAKDKTTERDLRTTPEPTTAAPKMTKEBAPTTPKE 751
Db      1306  SDEPTPSDEPTPEEPTPEEPIPTDTPSDEPTPSD---EPTPSDEPTPSDEPTPSDEPTPS 1361
QY      752  SKTPTTTPQVSTTQDTPPEKITLKTTPTLAPKVTTPKTTTTEINAKPEETAKPKDR 811
Db      1362  DEPTPSDEPTPSDEPTPEEPT---TTTPPTPSTT-----PTSG 1396
QY      812  ATNSKAT-----TPKPKPTKAP-KKPTSTKKPKTTPRVRKPTTTPPKM----- 856
Db      1397  SGGSSGSGGGGGGGGCTVPTSPPTPSKAPTAP--TELEPTPSDVPALIGEHNH 1453
QY      857  -----TSMELNPTSRIAEAMQTTPRQGTNSKLVEYNPKSEDGAGEETPHN- 908
Db      1454  YLRGPTDGSFPERNITRAEAIVF-----AKLL-----GADESYGQASASPSD 1498
QY      909  LLRPH-----VEMPE-----VPPDMOYLPRVPOGT----- 934
Db      1499  LADTHMAAMAIKFAATSGGLFKGYPGCTTKPPDONITRAEFAIYVLFLLKVKQEIWSKLA 1558
QY      935  ---TINPLASDETNICNG---KPYDGLTTL-----RNGTLVAFRGHYFWMLSPSPSPS 981
Db      1559  TIDISNPKFD---CVGHMAOEIEKLTLSGIYIGYDGT-----FKPON 1600
QY      982  PARITWEAGIPSPIDTFTRCNCEGKTFPKD--SQWRRTINDIKD 1026
Db      1601  YIKSESV-----ALINRALERGPLNGAKLPDPNNESTWAF-GDITMD 1642

```

RESULT 3  
T31108  
cyst germination specific acidic repeat protein precursor - *Phytophthora infestans*  
C:Species: *Phytophthora infestans* (potato late blight agent)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
R:Accession: T31108  
R:Goernhardt, B.  
submitted to the EMBL Data Library, April 1998  
A:Reference number: Z20986  
A:Accession: T31108  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-11489 <GOE>  
A:Cross-references: EMBL:AF061185, NID:q3851513, PID:q3851514, PIDN:AACT2308.1

C;Genetics:  
A;Gene: car90

Query Match	15.0%;	Score 853;	DB 2;	Length 1489;
Best Local Similarity	32.3%;	Pred. No. 5.3e-30;		
Matches 355;	Conservative 46;	Mismatches 475;	Indels 222;	Gaps 49;

Qy	9	TKKPRKPVNDVDEAGSGJDNDEKAVTTPDSTI-----QHAKVSTSPKITTAKPILNPR	62
Db	338	TPAPLEKYDYEBETTYTTEESTYAPVPSENNAPTERMHVAHIEKPCDTEVTYVAPTEET	397
Qy	63	PSLP-----PNSDT-----SKETSILVKKETTVEIKENTY---TNKQTSND	100
Db	398	TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYTPPEETTYAPTEETTYAPTEETTYAP	457
Qy	101	GKEKTTSAKETOSIEKTSAKDLAPTSKVLAPPKAEITTTKGALTPKBEPTPKBEPA	160
Db	458	TEETTYAPTEETTYAPTEETTYAPTEETTYAPV---BEETTYASPEETTYAPTEETTYA	514
Qy	161	STPKPEPTTTIKSAPTPPKBAPT--TKSAPTPPKBAPT--TYKBPATPKBAPT	217
Db	515	EETPEPEEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPEPEEET	573
Qy	218	TKBPATTTKSAPT-----TPKBPATPKBAPTTPKBPATPKBAPTTPKBE	267
Db	574	TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPEPEETTYAPTEETTYA	632
Qy	268	P-----AP---TKBPATTPKBP-----APAPKBKBPATPKBAPTTP	304
Db	633	PTTEETTYASTEETTYAPTEETTYAPAEETPEPEETTYAPTEETTYAPTEETTYAPTE	692
Qy	305	KBPAPT--TKKESPPTPKBP-----APT--TKKSAPT-----TKBP-----A	339
Db	693	TYAPTEETTYAPAEETPEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA	752
Qy	340	PT--TKKSAPT--TKKESPPTPKBP-----APTPKBPATPKBAPTTPKBPAPT--	388
Db	753	PTAEUTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	812
Qy	388	TPKBPATTTKBPATPKBAPTPPKBAPTTPPKLTPTPPKLAPTPPKBAPTPEE	448
Db	813	TPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET	870
Qy	449	LAPT--TEEPEPTTPPEP-----APT-----TPKAAPTPKBPAPTTPKBP-	489
Db	871	YAPTEETPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	930
Qy	490	----APTTPKBPATPKBAPTTPKBPAPTTPKBPAPT--TKKBPATPKBAPK--ELAPT--TT	539
Db	931	EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPEPEETTYAPTEET	990
Qy	540	KEPSTSDKRAPTPTPKGATTPPKBAPTTPKBP-----APTPPKGAPT-----	585
Db	991	YAPTEETMYAPTEETTYAPTEETTYAPAEETPEPEETTYAPTEETTYAPTEETTYA	1050
Qy	586	-----TLKBPATTPKBP-----APK-----LAPT--TTKGPTSDKRAPTP	624
Db	1051	EETTYAPTEETTYAPAEETPEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET	1110
Qy	625	KEPAPTTPKBPAPTTPKBPAPTTPKBPPTTSVSTPTTKBPT--TIKSPDSTPLS	682
Db	1111	YAPAEETPEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP	1169
Qy	683	AEPPPKA--LENSKREB-----VPTTKTAAPKPEMTTAAKOTTEROLRTPP---ETT	732
Db	1170	TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	1229
Qy	733	TAAPKAKKETATTEKTESKATATTOYSTTPTT-----PKITT---LKT	779
Db	1230	TYAP-----TEETTYPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	1284
Qy	780	TTLAPKVTTKTITTEIMNKPDETKAKDRAT--NSKATPKP-QAPTKAPKPT--	833

```

Db 1285 TTAAPLEATTATPEETPYAPTEETPEETGTYAPTEETTYAPTEETTYAPTEETPYE 1344
QY 834 -----STKRP-KTMRVRKKKPTTPPKMTSTMPELNTPSKRAEMLQTTTRP--N 881
Db 1345 PAESTSTVSTKRCPTTEFTDEPTDEPTDE--PSDEPTDEPTDEPTDEPTDEPTCDN 1402
QY 882 QPNSTKLVENPKSEDAG 899
Db 1403 QGINGIGVENKRYNNAG 1420

RESULT 4
116251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leimbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AA852641.1; GSPDB:GN00
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP.F35A5.1
A:Map position: X
A:Introns: 1272/2

```

Query Match 14.1%; Score 802; DB 2; Length 1274;  
 Best Local Similarity 28.0%; Pred. No. 7.2e-28;  
 Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

```

QY 6 KNRTRK-KP-----TP-----KPPVVDAGSGLDNGD-----FKYTTDTSTQHNKSTYS 50
Db 274 KNPKKMKPWEDETPVEEVKEPPVKAPVLKKKPAPAKADSPSKAAPKVEPS 333
QY 51 PKITTAPRI-NRPSLPNSTSTSETSJLVNKETVETKETTNNKOTSTDGKKTSAK 109
Db 334 SPVVPPEPVKPKPVKKRPVEVDEPRAEVKKPAPEKKTIVLKKKEPESSTTPSSDPS 393
QY 110 ETOSIEKTSANDLAPTSVLAKPTPKAE-----TTTKGPA-----LTPP 148
Db 394 PKKAPAPAKPRDSSPKKATPLQADPKAOEVPPTVKNPKVKKYKPPWEVDEDPVEEVKOP 453
QY 149 KEPTPTTP-----KEPASTTPKEP-----TP-TTISAPPTTPKEPAPTTTKSAPTTPKEPAP 199
Db 454 EAPAKKPTVLKRKEPAANDAKPATSKPTPEPEKKDPVKPPDSSPKVAAKPDSAQAPB 512
QY 200 TTTKEPA-----PTTPKEPAPTTTKEPAPT-----TTKSAPTTP 233
Db 513 TPVKKPVKKMRPWEDETPADVDVSKPTDAKKTSLAKKDPAPAKESLKPADTKAPAPK 572
QY 234 KEP-----APTTPKKPAPPTTPKE-----PAPTTPKEPTTTTKEPAPTTTKEPAPTTPK-EP 283
Db 573 RDPSPKVAAPTAPAEKKTIVLAKKEPAGPADSKTKEPESKRPKRAVPKAPVKEV 632
QY 284 APTAKKKAP-----TTPKEPAPTTTPKEPAPTTPKEP-----SPTTPKEPAPTTTKSAPT 333
Db 633 AAAAAYKKBEPSKPKDLPKKAKEPSPVVP-PTPKKNVKKMKPPWEDDADAPAKVSLPE 691
QY 334 TTKEPAPTTTKSAPTTP-KEPSPTTTKEPAPTTPK-EPAPTTPKKAP--TTTKEPAPT 388
Db 692 PEKK-TPVLAKKAPTKPDESEAAADVSGSPSKDKPLAKKAPVKKPRDPSKAPVPIKAPK 750
QY 389 TKREPAPTTTTPKAPPA-PKEPAPTTTPKETATTPTKKLLPTTP-----EKLAPTTPKEPA 442
Db 751 T--EVPAPAVVKKPEVAKSRDPSKKAK-AEENSP--VVPPTPVKNPVKKMKPPWEDDA 805

```

```

QY 443 PTTPEELAPTTPEED-----TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPT--PKE 496
Db 806 PAEPVNVPEPEPKKTIVLAKKTPVKRDPSPKRAVPAPKSTKTDAAPVSVKKPEPVSKPE 865
QY 437 PAPTTPKETAPPTPKGTAAPTTLKEPAPTTPKKPAKELAPT-TTKEPTSTSDKPAPTTP 555
Db 866 PSEKKAEPNSPVVP-----PTPVKNPVKKK-KPPMEDDEPEIEVKKKSE--PEKKTIVLA 918
QY 556 KCTAPTTKEPAPTTTPKEPAPTTPKGTAAPTTLKEBAPT-----TTKKKAP-----KEL 603
Db 919 K-KEPEKPRD-APKVAARKRDPSPKKAYPE--KEPAKVAARKRDLSPKALPIPANTQEA 974
QY 604 APTTTKGPTIS-----TSDKP-----APTTPKET-----APTTPKEPAPTTPKKAPPT 646
Db 975 PPTPVKNPVKKMKPPWEDEDEPAEVSAPPEPKKTIVLAKKAPAKPRP--SKKAPAV 1031
QY 647 TPEPTPTTSEVSTPTTKEPTTIKSP-----DESTPELSA-EP--TPKALENSPKRPG 698
Db 1032 AAK-FDPKRIPEV-PTTPVKNPVKKMKPPWEDEDESESVASPEPEKKTIVLAKKAPTKPA 1089
QY 699 V-----PTTKTPAAT-----KPEMTTAKDKTERDLRTTPTT--TAAPK 737
Db 1090 TKPDEAAADPVSGPTSKDPLSKKAPVEKRPPTTDPKDKLKPSPAKPEKAPPAAPK 1149
QY 738 MKKETATTTKTESKITATTOVSTTQDTTPPKITTLKTTLAPVTTTKTITTE 797
Db 1150 KMKPVWDDDPDEPADFTVPAPSKKPDIEDPDLG-----GPKTKDPK----- 1193
QY 798 INNKPEETAKPKDRATNSKATTPKPKOKPTKAPKKTSTTKKPTTPRVKPK----- 848
Db 1194 -LNKKAPAKKPTKE-----PKREVSKPEPKPTPEPKP-AAPKKMKPPWEDDPDEPE 1243
QY 849 ---TTPTPKMTSTMP 861
Db 1244 ADEFTMPAPKKPDIEDP 1259

```

RESULT 5  
 S48478  
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 12-Nov-1999  
 C:Accession: S48478; A26877; S27281; JG6123  
 R:Rowley, K.  
 submitted to the EMBL Data Library, October 1994  
 A:Reference number: S48478  
 A:Accession: S48478  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <ROW>  
 A:Cross-references: GB:247047; EMBL:Z38061; NID:9603997; PID:g763364; GSPDB:GN00009;  
 J. Bacteriol. 169, 2142-2149, 1987  
 R.Yamashita, I.; Nakamura, M.; Fukui, S.  
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STR1.  
 A:Reference number: A91831; M0ID:87194600  
 A:Accession: A26877  
 A:Residues: 1-242 <YAM>  
 A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525  
 A:Accession: B26877  
 A:Molecule type: DNA  
 A:Residues: 762-1331 <YAZ>  
 A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526  
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
 FEBS Lett. 239, 179-184, 1988  
 A:Title: Similar short elements in the 5' regions of the STR2 and SGA genes from Sacc  
 A:Reference number: S27281; M0ID:89031230  
 A:Accession: S27281  
 A:Molecule type: DNA  
 A:Residues: 1-31 <PAG>  
 A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552  
 R.Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.



[illegible]

RESULT 7  
T03826  
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse  
N:Alternate names: alpha-NAC protein -  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 24-Nov-1999  
C:Accession: T30826  
R:Yotlov, W.V.; St-Arnaud, R.  
Genes Dev. 10, 1765-1772, 1996  
A>Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle  
A:Reference number: Z20889; MUID:96312450  
A:Accession: T30826  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2187 <YOT>  
A:Cross-references: EMBL:U48363; NID:q1666688; PID:q1666689; PIDN:AA818732.1  
C:Genetics:  
A:Gene: Naca  
A:Map position: 10  
A:Introns: 24/1, 1996/1, 2024/3, 2050/3, 2099/3, 2142/3, 2183/3  
A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding acti  
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match	13.6%;	Score 776;	DB 2;	Length 2187;
Best Local Similarity	27.7%;	Pred. No. 1.6e-26;		
Matches 288;	Conservative 122;	Mismatches 389;	Indels 240;	Gaps 49;

```

Oy      34  VTTPDSTJOTONHNKSTSPKTPYTKAPINPSPSLPSPNSDSKE-----TSLVNNKETT 84
Db      806  VQSKRVPIMSDYTPSPSKTSATAV-PK-----DTSATLSKSPAVTSLSPKAPV 857
Oy      85  VETKETTITTKQSTDOCKEKTTSAKETOSLEKTSAMDIAPIKVLAKTTPRAEITTKGPA 144
Db      858  APSNEATIVPEITPSLKNAIAATKREILANSIPKVTSPS-----PQTPKPSVSLKAGPA 913
Oy      145  LTT-----PKK-----PIPTTPKPEASTT-----PKREPIPTT 171
Db      914  MTSKATEIAASKDVSPQCPPEPLIOHVPTSPKSVSDTSLGALTSPPKOP-PAT 972
Oy      112  IKSAPTPPEKE-PATTPYTKKSAPTPK-----EPAPYTKKAPATPKKAPATTKEP 221
Db      973  IAEPTTYTKKSPKAASKKTPATPSPBGVYAVPLELIPQSKAKAPTAAPKESASSTSR 1032
Oy      222  APPTT--TKSAPT-----TKKPEAPYTPPK-PATYTPKKEAPYTP 257
Db      1033  APTKIVNSKEIKSPGVAVPLEISLPIKETSASATPEKSAKSPKSPKTAGKE--TPP 1089
Oy      258  KPPYTPYTKKEAPYTPKKEAPYTPKKEAPYTPKPKKAPATTT--PKKPAV---TTPKKEAPYTP 311

```

```

Db      1090 GGVAVAPPEISLPKPETPONATPNESI,LAASSQKKSPTISVPEKPEPGGVATAMLEIPSA 1149
Qy      312 TKEPSPTPK-----EPAPTTTKSAPTTTKEBAPT--KSAPTTPK-----EPSPTT 357
Db      1150 OKAPRTAVPQOIPPEDAVITLLAGSPLSPKKASXTAAPKEAPAPPSGVIAVSGEISPS 1209
Qy      368 TKEPAPTPKPEAPITPKKAPPT--PKE-----PAPTTKEPAP-----T 396
Db      1210 KTSKTAAPKENSATLPPKRSPTAPAREVPAISSEGVAVSPSISPSPTPASKGVPT 1269
Qy      397 TTKKAPAPAKPEPAPPTTKEBAPTTPKKLPTTEKLAPTTPEK-PAPTTPEELAPTTPE 1455
Db      1270 LTPKAPAMLALE-SPASPKVAVKVTAAPEETSTP-----SPQIKIPVAGKEKASAPPS 1322
Qy      456 EPTPTT-----PEEBAPTPPKAAANTKEBAPTTPKE-PAPTTPKKEBAPTTPKETA 506
Db      1323 KKTPTAVPKEITSASEGVAVLEIPSPSPKAPKTAAPKEBTPAPS--PEGATTAPOIP 1380
Qy      507 PTTPGTATTTTKEBAPTTP-----KKAPKELAPTTTKEPTS-- 544
Db      1381 PEPKRGSKKAGKE-TPTTSPSEGVTAAPLEIPSSKTKSMASKEPLVPPSSKKLSOT 1439
Qy      545 ----TTSOKPAPTTKGVAPTTPKPEBAPTTPKE-PAPTTPGTATTTTKEBAPTTPKPA 599
Db      1440 VGPKEITSLEGAIVALEIPSSHKAAPTVDPOVLTSPK-DAPTTLAE-SPSSPK-A 1496
Qy      600 KPELAPTTTKGPTSTSDSOKPAPTTKETAPTTPKPEBAPTTPKKAAPTTPETP-----PPT 654
Db      1497 PKTAAFPSR-VTYVPEKPA-TPOKASGTASKVPVAEQEVAVSRETPVTPAVPV 1554
Qy      655 TSEVSTPTTTK-----EPTTIHKSPEDESTP,LSAEPTPKALENSPKE-PCVPTTKPAA 707
Db      1555 KNPSSHKRSKTEILEKAPATLPPSTPSKPISSKKAAPT--SAPKEFPASPSTK-PVT 1611
Qy      708 TKPEMTTAKDKTTERDLPPTPETTAAPKMTKETATTTEKTESEKITATTTQVSTTQ 767
Db      1612 T-SLQOTAPPSQKAPSTTIKREMLAPV----LVPSSKSPAPAPARASASLSPATAAP 1665
Qy      768 DTPPEKITLKTTLTAAKVYTTKKTITTTTETINMKKEEELAKPPDRATNSKATTPKQKPTK 827
Db      1666 QUPAKKATITPSCKKAARETEPLETSTAPLSGAKRETSE--TSSVKVLMSSP----- 1716
Qy      828 APAPKPTSTKPKMPAVRKPPTTPPRKMTSMPELNPSTRIAEAMLOTTPRPNQTPNSK 887
Db      1717 -PKKASSSKRASLLP-----ATLPLSKASVLS-----PTATSSGK 1752

```

RESULT 8  
A:35175  
Mucin 1 precursor, repetitive splice form A [validated] - human  
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin  
N:ncratic mucin; polymorphic epithelial mucin (PEM)  
C:Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor  
C:Species: Homo sapiens (man)  
C:Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000  
C:Accession: A35175; A35175; A35866; A35867; S10572; S40293; A36735; PX0066; S10218;  
R:Litgenberg, M.J.L.; Vos, H.L.; Gemmissen, A.M.C.; Hilkens, J.  
J. Biol. Chem. 265, 5573-5578, 1990  
A:Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene  
A:Reference number: A35175; MUID:90202794  
A:Accession: A35175  
A:Molecule type: mRNA  
A:Residues: 1-952,1033-1344 <LI61>  
A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124;  
A:Experimental source: splice form A  
A:Note: Genbank entries HUMEPIS1A1 and HUMEPIS1A2 present only the amino-and carboxyl  
A:Accession: B35175  
A:Molecule type: mRNA  
A:Residues: 1-19, 29-952,1033-1344 <LI62>

A:Cross-references: GB:M32739; GB:J05288; NID:9182126; PIDN:AAA55806.1; PID:9182129; GB:  
A:Experimental source: splice form B  
A:Note: GenBank entries HUMPEP1 and HUMPEP1B2 present only the amino- and carboxyl-ter  
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Pedemonte, J.; Dunlap, T.; Peat, N.; Burchell  
J. Biol. Chem. 265, 15286-15293, 1990  
A:Title: Molecular cloning and expression of human tumor-associated polymorphic epithel  
A:Reference number: A35886; MUID:90368715  
A:Accession: A35886  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-19,29-992,1033-1344 <GEN>  
A:Cross-references: GB:J05581; NID:9188669; PIDN:AAA59876.1; PID:9188870  
A:Note: GenBank entry HUMWUCAB includes one copy of the tandemly repeated sequence  
J. Biol. Chem. 265, 15294-15299, 1990  
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.  
A:Reference number: A35887; MUID:90368716  
A:Accession: A35887  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>  
A:Cross-references: GB:J05582; NID:9189598; PIDN:AAA60019.1; PID:9189599  
A:Note: GenBank entry HUMPANUW contains four fewer copies of the tandemly repeated sequ  
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.  
Eur. J. Biochem. 189, 463-473, 1990  
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera  
A:Reference number: S10571; MUID:90276413  
A:Accession: S10572  
A:Molecule type: mRNA  
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WR2  
A:Cross-references: EMBL:X52229; NID:937053; PIDN:CAA36478.1; PID:937054  
R:Abel, M.; Sidiqul, J.; Kufe, D.  
Biochem. Biophys. Res. Commun. 165, 644-649, 1989  
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated  
A:Reference number: A36735; MUID:90088473  
A:Accession: A36735  
A:Molecule type: mRNA  
A:Residues: 1-142,'Q',144-162,'Q',164-168 <AB>  
A:Cross-references: EMBL:M31823; NID:9181542; PIDN:AAA35757.1; PID:9181543  
R:Masuzawa, Y.; Myauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.  
J. Biochem. 112, 609-615, 1992  
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu  
A:Reference number: JX0235; MUID:93123189  
A:Accession: PX0066  
A:Molecule type: mRNA  
A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>  
A:Experimental source: gastric carcinoma cell  
R:Zirhan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.  
FEBS Lett. 356, 130-136, 1994  
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r  
A:Reference number: S51026; MUID:95080414  
A:Accession: S51026  
A:Contents: annotation  
A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region at  
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c  
partial repeats. The repeat shown is defined by SmaI nuclease sites.  
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly  
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146  
A:Genetics: GDB:MUC1; PUM  
A:Cross-references: GDB:120705; OMIM:158340  
A:Map position: 1q21-1q23  
A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3  
C:Superfamily: polymorphic epithelial mucin  
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis  
F:1-1344/Product: mucin I precursor, splice form A #status predicted <PREA>  
F:1-62/Region: mucin I amino-terminal non-repetitive

F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>  
F:1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>  
F:1-19,29-1344/Product: mucin I precursor, splice form B #status predicted <PREB>  
F:1-19,29-212,1033-1344/Product: mucin I precursor, epithelial tumor antigen splice f  
F:138-1017/Region: 20-residue repeats (GSPAPAHGVTSAPDRPAP)  
F:1143-1344/Region: mucin I carboxyl-terminal non-repetitive  
F:1245-1272/Domain: transmembrane #status predicted <TRM>  
F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predic  
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 11.8%; Score 672; DB 1; Length 1344;  
Best Local Similarity 27.8%; Pred. No. 3,1e-22;  
Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;

QY 77 LVNKKETVEKETTNTNK--QTSQGEKETSAREKOSIEKTSKD-LAPTSKVLAKPT 133  
DB 15 LVNLTATTAPKPAVTVTSGHASSVTPGGEKETSATQSRSSVSSTKKNVSMVSSVLSHS 74  
QY 134 P-KAETTKGP--ALTTRKEP-----TPTRKEPASTTP-----KE 166  
DB 75 PEGSSSTTQGDVTLAPATEPAGSAAATMGODVTSVPYTRPALGSTRPAAHDVTSAPDNK 134  
QY 167 PPTPT-----IKSAPTPKKEPAPTT-----KSAPTPKKEPAPTTKEP-----A 206  
DB 135 PAPGSTAPPAHGVTSAPDT--RPAGGSTAPPAHGVTSAPDT--RPAGGSTAPPAHGVTS 190  
QY 207 PTPKKEPAPTTKKEPAPTTKSAPTTPKKEPAPTP-----KKPAPTPKEP----- 252  
DB 191 PDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGV 247  
QY 253 --APTPKKEPPTTP-----KEPAPTTKEP-----APTPKKEPAPTPK----- 289  
DB 248 TSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGV 307  
QY 290 KPAPTPKKEPAPTPK-----EPAPTTKKEPPTTPKKEPAPTT-----TSAPPTKKEPAPTT 342  
DB 308 TSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGV 367  
QY 343 TKSAPPTKKEPPTTPK-----APTPKKEPAPTPK-----KPAPTPKKEPAPTPK----- 391  
DB 368 T-SAPDT--RPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPA 424  
QY 392 --EPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTPE 447  
DB 425 HGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPA 484  
QY 448 --ELAPTPPEEPPTTP-----EEBPAP--TTPKA-----AABNTPKKEPAPTPK----- 487  
DB 485 HGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPA 544  
QY 488 --EPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTT-----KKPAPKEL 534  
DB 545 HGVTSAPDTRAPAGSTAP--PAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGST 600  
QY 535 APPTTKKEPPTTSDKPAAP--TTPKGTAPTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTT 588  
DB 601 AP-PAHGVTSAPDTRAPAGSTAPPAHGVTS-----APDTRAPAGSTAPPAHGVTSAPDTR 655  
QY 589 EPAPTTTP-----KKPAPKELAPTTKKEPPTTSDKPAAP--TTPK-----ETAPTPP 632  
DB 656 APGSTAPPAHGVTSAPDTRAPAGSTAP--PAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTR 714  
QY 633 KEAPPTTP-----KKPAPTPPTTPPTTSEVSTPTTKEPTT-----IKSPD 675  
DB 715 PAPGSTAPPAHGVTSAPDTRAPAGS--TAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPD 772  
QY 676 ESPPELSAEPTRPALENSPKPEGVPTTKTPAA-----TKPEMTTJAKDK---TTERD 724  
DB 773 TRPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPD 832  
QY 725 LKTPPTT-----TAAKMTKETATTEKTESKITATTTQVSTTTQDPTT--FKIT 775

Db 833 TRPAGSTAPPAHGVTSAP----DTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGHVT 888  
Qy 776 TLTETTLAPKVTITTKTITTT--EIMNKPEETAKPKDRATNSKATTPKPKOKTKAPKKPT 833  
Db 889 SAPDTRPAGSTAPPAGHVTSA PDTRPAGSTAPPAGHVTSA PDTRPAP--GSTAPPAHG 946  
Qy 834 STKKPKTPRVKPKPTTPPKMTSTMPELN---TSRIEAMLOI--TTRP---NQTPN 885  
Db 947 VTSAPDTRP---APGSTAPPAGHVTSA PDTRPAGSTAPPAGHVTSA PDTRPAGSTAPP 1003  
Qy 886 SKLVEKNPSEDAAGAEGETPHMLRPHVMEVTPDMOYLPRVNOGIIINPMLSDEFN 945  
Db 1004 AHGVTSAPDTRPAGSTAPPAPH-----GVTSAPDNKRALGSTA---PPVHNVTIS 1049  
Qy 946 ICNGKPVDDLTLRNGTLVAFRGHYFMWLSPPSPPS 981  
Db 1050 ASGSASGSASTLVHNGTSARATTTAPSKSTPSPIS 1085

RESULT 9  
T18535  
high molecular mass nuclear antigen - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18535  
R:Shimada, K.; Harata, M.; Mizuno, S.  
J. Cell Sci. 110, 3031-3041, 1997  
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick  
A:Reference number: Z18955; MUID:9803440  
A:Accession: T18535  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1151 <SH1>  
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BA24137.1

Query Match 11.8%; Score 671; DB 2; Length 1151;  
Best Local Similarity 25.9%; Pred. No. 3e-22;  
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;  
Qy 14 TPKP-----PYVDAGSLDNGDFKVTTPDTSTQHNKVSYS--PKITTAKPINRRS 64  
Db 113 TTPPSQGPAGTTPPSQGAAGAPKGDGTAGPSGTGSGADGKPAQDVYKATTA--ATEARP- 170  
Qy 65 LPPKSDTSKETSULVKNKETEVEFKETTTNK--OTSQDGEKETSKEKOSIEKTSAKDL 122  
Db 171 -----ASAASPTVPKATAEATAVTAASQASAPKAADAAVTA--SOSAKATV-EV 219  
Qy 123 APTSKVLAKPTPKAETTTKGPALTTPKEPTTP--TPPKEPASTTPKEP---TPTTTSAPT 177  
Db 220 KPAAAVAKAEKAVTAATAAAKPAATAEAKPAPVTSPTIPCSAEAKPLTAASPTASKA--T 277  
Qy 178 TPKEPAPT-----TKSAPTTPKEPAPTTPK----- 203  
Db 278 AEAKPVATASLMAKTATAEKAPAPSPSVKATTDITKAATATAPKAGPDVKAIVAVCAEA 337  
Qy 204 EPATTTKEPAPTTPKESAPT-----PKEP-----PTTKKPAPTTPK 250  
Db 338 KPAPPPPOOLPKAAAAAATPGTLKATATAPPHCSPPRANHVTIVTPPNVPRAAAATVP- 396  
Qy 251 EPAPTTPKEPTPTT---PKEPAPTTPKEPAPTTPKAPPAKAPAPTTPKEPAPTTPKE 306  
Db 397 -TAGAVPKASTGTTPAAPAOQVPV---KAAVTPTPSPQOAVPRAATAA--APVTPQO 448  
Qy 307 P---APTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKE 360  
Db 449 PVTKAATTTNATPPQOLPKAATTTTATPVTPQOIPKAGTDAAPPAVPAASDGAAT 508  
Qy 361 P---APTTPKEPAPTTPKAPAPTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTPKE 406  
Db 509 PGVPAATADPOKPPPTPOSVSATTEKPKQRAAPPSNENATPAVSPSPKSLPLPTIP 568  
Qy 407 K-----EPAPTTPK---ETAPTTTPKKLPTPEKLAPTTPKEPAPTTPPELA----- 450

Db 569 KPVPLMALITQPVTAQWVQLAATKRPSPVYKASPK-ALMTPEPPPGGLPRALAAAKLLG 627  
Qy 451 -PTTP-----EETPTTP---EEPAPTTP-----KAAAPTTPKE----- 480  
Db 628 LPSSPVSAMHAKVTPPLPASPVPMAASPSLCPDAAVLAATNAASPAKPPAAGNG 687  
Qy 481 -PAPTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKE 516  
Db 688 TLMAVMGANTOMAPIGAGAQAOTAPMGAAHTHVSPPMGAGATQMSGTGAANTHMSPIGA 747  
Qy 517 -----TLKEPAPTTPKPPAPKELAPTTTKEP-----TSTTSKPPAPTTP 555  
Db 748 GGATQMSPMGAAQNTQMSMGATTTQMSMGAAATQSPMGAAATQYATISAGNTMQVSP 807  
Qy 556 KG--TAPTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKE 611  
Db 808 MGAATPPQTPSVGAATTP--QSPPM---GAATVTLMSPGCAATTPQ---PSPMGAVTTQPP 859  
Qy 612 -TSTTSKPPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKE 652  
Db 860 PMAATNTTQPPMAASTQTPMGAAATTTQSPMGAAATTTQSPMGAAATTTQSPMGAAATTTQSP 919  
Qy 653 PVTSEVSTPTTKEPTTIHKSPPDESTPEL--SAEPTPKALENSKREQVPTTKTPATKP 710  
Db 920 PPPPIPPSPATQSPQPMKSPDPDPKAPSAQAOTSPAHHVANAASPCV--TAVSPA---P 975  
Qy 711 EMTTAKKKTERTDRTPEPTTTAPKMT--KETATTEKTESKTTATTTQVSTTTQDT 769  
Db 976 IGVTAESPADGARSPOGTLAATDQPKASPAATADVTAAID--VTAAATVAPA-----EA 1029  
Qy 770 TPKKITTLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPKPKOKPTKAP 829  
Db 1030 AP-----TKAKRSS 1077  
Qy 830 KKPSTKKKTKIMPRVRKKTTPTPPKMTSTMPELNPTSRIEAMLOITTRNQPNKSLV 889  
Db 1078 -----PAVGDDGQOQNTPGAQSVPP-----VTEAAVQ----- 1104  
Qy 890 EYNPKSEDGAGE 904  
Db 1105 EAAAAAAGAGARE 1119

RESULT 10  
T25697  
hypothetical protein F16F9.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T25697  
R:Fulton, B.  
submitted to the EMBL Data Library, August 1996  
A:Description: The sequence of C. elegans cosmid F16F9.  
A:Reference number: Z20071  
A:Accession: T25697  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1229 <FULL>  
A:Cross-references: EMBL:067956; PIDN:AA07691.1; GSPDB:GN00028; CESP:F16F9.2  
A:Experimental source: strain Bristol N2; clone F16F9  
C:Genetics:  
A:Gene: CESP:F16F9.2  
A:Map position: X  
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 11.6%; Score 659.5; DB 2; Length 1229;  
Best Local Similarity 29.0%; Pred. No. 1e-21;  
Matches 307; Conservative 78; Mismatches 339; Indels 333; Gaps 52;  
Qy 33 KVTI-PDTSITVO---HNKVSIPKTTTA-KPINRPSLPPNSDT-----SKETSILYN 80  
Db 119 KVTISDASITNAPJTQKDSITPEILITGIVINSKSESVTDMSTWRFSTLLSPTELLNS 178



```

QY 81 KETTVEKETTNNKQSTDKERTTSAKENQSIKTSKADLAPTSKVLAKPTPKAETTT 140
D 179 PETLVSDSSSTST-EQTSPODNTIELASPMENINTTEATTSSVESSVGLAS---EDETTV 234
QY 141 KGPLATTPKEPTPTTPKEPAATTPKEPTPTTKSAPPTPKAPATTPPKSAPPTPKAPAT 200
D 235 TALEST-----TTLVAESTTTEETPTT-----ASTTKKSTT-----KAPA 272
QY 201 TTKEPAATTPKEPAATTPPKSAPPTPKAPATTPPKAPATTPPKAPATTPPKAPAT 256
D 273 TTEPTPTTTEE--VTTEATSTTTSSSESTEK---PTPLDINKTAGPATG---PET 324
QY 257 PKEP-TPTTPKEPAATTPKEPAATTPPKAPATPKAPATTPPKAPATTPPKAPAT 314
D 325 THFVGTGTTTP--FDIATEPFAKSEDKMTLSKTATETTTQOTTEVT-DGPEKETTKV 361
QY 315 ----PSPTTPKEPAATTPKSA-----TTKEPAAT-----TKSAPTPPKESPTTK 359
D 382 SIEPTTTPVLVETSTSTTSKESDGFHTTLKLVTTADSDSTESATVTKPNEETTK 441
QY 360 E---PAPT-----TPK-----EPAPTTPKKAP-----TTPK 383
D 442 SHVPRPTKKGTVKVTPLKELSDDEPTEIT-KAPHPCKLLEKTYHPVLSDNARISEAK 500
QY 384 E-----PAPTTPKEPA-----TTTKKAPATA--PKEPAPT---PKETAP 419
D 501 ENDQYNLDYNHVREAKEPTTTESSSTTEETVTTTEEPANGNPPTENPTTEQPTSTAE 560
QY 420 TTPKLLPTTPPEKLA---PTTPKEPAATTPPELAPTTPEPTPTTPEEPAPTTPKAAAP 476
D 561 STTALPPTTEQVVTTEPTTAESTANO---KPTTQOSVST---EKSTTKKA---S 610
QY 477 TPKEPAATTPKEPAATTPPKAPATTPPKGAPATTPKEPAATTPPKAPATTPKAP 536
D 611 TTEB--PTTDEPTTTT---ESSTGKATPBEISTSEETTELKITE-----GS 657
QY 537 TTKKEPTSTSDKAP---TTKGAAPTTPKAPATTPKAPATTPKAPATTPKAPATTPKAP 592
D 658 TTEETPTTAIFAEASTGIIITDEETSTSTPEITSTKE--IYESAITQISVSVES 715
QY 593 TTP-----KAPKAPATTPTKG-----TS 613
D 716 STPRLPERKKAIVNFKHLEVLAKKKLKEKSTSGDSSSTTTVAENIDEVT 775
QY 614 TTSCKAPATTPKEPAATTPKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPK 700
D 776 TEKEKVVQTTPTTEKSTQOETTTTTEKTSKTTTEKPTTSATTEITTSPT- 834
QY 671 HKSPDSTBELSAEPTPKALENSPKEPVPTTKPATPEMTTAKDKTTERDLTPPE 730
D 835 ----TEST-----TVDTSATTEESSTAETTTSAE---TSE 865
QY 731 TTTA-----APMTKETTTTEKTE 751
D 866 TTTSSEAAFITGSPENTALOSSSOKSENESSAEKRGARDFVKKAKHTTVKPAETTTA 925
QY 752 SKTATATTO-VIST---TQODTTPKITTLLKTTTLAPKVT---TTKKTITTEINKBEE 804
D 926 VAASITTTPTTTEKSTLTETTPTEATTLNENVTGAPFVGTGAPVDETITNTELLSK--- 982
QY 805 TAKPKDRATNSKATTPKPO-----KTPKAKKCPSTKPKPTMRRVKKP 847
D 983 -----INNTQISQKPRPIDISKTDALSLIGSLGFTKAMAPTIT----- 1022
QY 848 KTTTPPKKMTSTMPELN-----PTSRIAE 872
D 1023 HTTDAAFVATATEASLNDGSDKKIIDEAQPTEIRRA 1059

```

RESULT 11  
T11622  
extensin class 1 precursor - cowpea

C:Species: Vigna unguiculata (cowpea)  
C:Date: 16-Jul-1999 #sequence-revision 16-Jul-1999 #text-change 17-Nov-2000  
C:Accession: T11622; S54135  
R:Arsemlievic-Maksimovic, I.; Broughton, W.J.; Krause, A.  
Mol. Plant Microbe Interact. 10, 95-101, 1997  
A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.  
A:Reference number: 217301; MUID:97155574  
A:Accession: T11622  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-489 <ARS>  
A:Cross-references: EMBL:X91836; NID:971015936; PIDN:CA62943.1; PID:971015937  
A:Experimental source: sub-species Red calcoana  
R:Arsemlievic-Maksimovic, I.; Broughton, W.J.; Krause, A.  
Submitted to the EMBL Data Library, April 1995  
A:Description: A class of root-hair specific extensins involved in rhizobium/legume  
A:Reference number: S54155  
A:Accession: S54155  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 326-489 <AR2>  
A:Cross-references: EMBL:X86030; NID:97911449; PID:9791150  
C:Genetics:  
A:Gene: Ext26g  
C:Superfamily: hydroxyproline-rich glycoprotein  
C:Keywords: glycoprotein; hydroxyproline  
F.1-23/Domain: signal sequence #status predicted <SIG>  
F.24-489/Product: extensin class 1 #status predicted <MAT>

Query Match 11.1%; Score 633; DB 2: Length 489;  
Best Local Similarity 31.2%; Pred. No. 5.7e-21;  
Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 7;

```

QY 179 PKEPATTPKSAPTTPKEPAATTPKEPAATTPKEPAATTPKSAPTTPKEPA 238
D 39 PKQTPPYVYNAAPPYVYKSPSP-----PSPSP-----PPYVAKYPPYVYKSPSP 83
QY 239 TTPKAPATTPKEPAATTPKEPTPTTPKEPAATTPKEPAATTPKAPATTPK 295
D 84 PSPSPPPYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 141
QY 296 PKEPATTPKAPATTPKEPTTPKEPAATTPKSAPTTPKEPAATTPKSAPTTPK 355
D 142 YKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 187
QY 356 TTPKAPATTPKEPAATTPKAPATTPKEPAATTPKEPAATTPKAPATTPKAPATTPK 415
D 188 YVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 247
QY 416 ETAPTPPKLLPTTPPEKLAPTTPKEPAATTPPELAPTTPEPTPTTPEEPAPTTP 475
D 248 PPPPYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 307
QY 476 NTPKEPAATTPKEPAATTPKAPATTPKAPATTPKGAAPTTLKEPAATTPKAPKELA 535
D 308 PSPSPPPYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 366
QY 536 PTTKEPTSTSDKAPATTPKGAAPTTPKEPAATTPKEPAATTPKGAAPTTLKEPAATTP 595
D 367 ----KSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 422
QY 596 KKPAPKELAPTTTGGPTSTSDKAPATTPKATPATTPKAPATTPKAPATTPPETPP 653
D 423 SP-----PPYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 475

```

RESULT 12  
I38346  
elastic titlin - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 29-May-1998 #sequence-revision 29-May-1998 #text-change 21-Jul-2000  
C:Accession: I38346

R:Labelt, S.; Kolmerer, B.  
Science 270, 293-296, 1995

A;Reference number: A57430; MUID:96026330

A;Accession: I38346

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-7502 <KEB>  
A;Cross-references: EMBI

C:Genetics:

A;Gene: GDB:TTN

A;C;Cross-references: GDB:127867; OMIM:188840

A;Map position: 2q31-2q31

Query Match	11.1%	Score 632.5;	DB 2;	Length 7962;
Best Local Similarity	22.9%	Pred. No. 9e-20;		
Rest 295; Conservative	105;	Mismatches 453;	Indels 437;	Gaps 60

```

QY 10 KKKPP- -KPPV- -VDEAGSLDNGCFKVT- - - - -TPDSTTQHNKVS 50
Db 5837 KKAAPPKVIKAPVIEKISTSRMEEEKVQVTIVKEVSKIVPOKSPSTPOVEEVK 5896
QY 51 -PKITTAIPINRPSLPRMSDTSKESL- - -VNKETVETKETTITNKOTSIDGKEKT 106
Db 5897 VAVVHTKKVNIISSEKMFASHTHEESVTVPEOKEIVTEKHVAVSKVE- - - - -P 5549
QY 107 SAKETOSIKETSAKIDLAP- -SKVLAKPPKAEITTKGALTTPKREPPTPKEPAS- - - 161
Db 5950 PKVPELPRKPAPEEVAVPIDPKVBPAPKVEVKKP- -VPEKKKVPVPKKEKPAAPK 6008
QY 162 - - -TTREPIPTTIKSAPTTPKEPAP- - - - -TTKSAPTPKKEAPITTEPA- 206
Db 6009 VPEVPKRPPEEKIPVAVAKKKEAPKAVPEOKGVVTEKITIIVTORESB- - - - -PPAV 6064
QY 207 PTPPKREPAPTTKEPAPITTKSAPTTPKKEPAPTPPKPAATTP- -KEPAPTPKEPPTT- 264
Db 6065 PELPKKVVDEER- -PVPRKEEVPKPPKVA- -LPKKVPEEKVAVPVPAKAPBRAE 6121
QY 265 - - - - -P 265
Db 6122 VSKKTVEEKRFVAEEKLSFAPVQVETVRHEVSADEMSYSSEEBGSISVYREEERE 6181
QY 266 KEPAPTT- - - - -EPAPTT- - - - -KPEA- - - - - 284
Db 6182 EEEAEVTEYVMEPREBEYVVEEKLIHSKRAVEAPRAVETROEKKIYLKXIKPAKIEBP 6241
QY 285 - - - - -PPAPK- - - - -PAPTPKE- -PAPTPKEP- -APTTKEPSTTPKE- - - - -PAPT 326
Db 6242 PAKVPAPKPIVEEKVAPVAPVPRKEKVPVPEEKPPVPEKVPKVIKMEEPRLAKV 6301
QY 327 TTKSAPTTTPKE- - -APTTTSAP- - - - -TPKEPSTTPKE- - - - - -PAPT- - - 365
Db 6302 TEKHMOITOEKVLAVATTKKAPPAKAVPDEBPRAVDEKVLKLPKREBPRAKVEFR 6361
QY 366 - - - - -PK-EPAP- - - - - 371
Db 6362 KRYVKEEKVSIAPKRPQOIKEVTIMEEKRAVTTLEEAVSVOREEEYEEEDYKEF 6421
QY 372 - - - - - - - - - - -TTPKKAPT- - -TKREPAPTPPKKEPAPTTK 399
Db 6422 EEEYPTTEYDOYEEREYERYEENEHYITEPKEPIPVKVPPEEVPVTKAPAPKAVLK 6481
QY 400 KPAPTAKREPAPTPKETAATTPKTLKLPPTPEK- - - - -LAPTPKPAATPTEELAPT 452
Db 6482 KAVP- - - - -LPKLIKPPPK- - -VPEEKKVTEEEKIHISIKRKKEQUTEP- - - - -AAK 6534
QY 453 TPPEPTPTTEEPAPTPKKAAP- - - - -NTPKEPAPTTKEPA- - - - - 490
Db 6535 VPMKPRVVAEKEKVPVPRKEVAPVPRVPEVKLEPEBEVAFEEBVVTHVEEYLVEEBEY 6594
QY 491 - - - - -PTTKREPAPTPKKEPAPTPKGAAPTTLKKEPAPTPPKKPAPELAPTTTKEPT 544

```

Db	6595	IIHEEEFTTEEEVVPVLPVK -VPEVPRKRPVEE -KKVPVPEKKKEAPPAKPEVPEKKDEE	6652
QY	545	TT-----SOKRAPPTKGGAPPTTPKBEAPPTPKBEAPATTTKGAPTLTKBPA--PTTPK	596
Db	6653	KVPVLPKPEKKR---PAKVAPEVPEKKVP--EEKVPVPEKVPKEAP-----PAKVPEVPK	6702
QY	597	KPAPEL-----APTTKGPTSTTSBKPAAPTPPKETAPT--TKPE---PAPTTPKKAPATTP	648
Db	6703	KPVEBKAPVPAPPKVPEAPPAKVPEPRPKLLIPEKKPTTPVPEKVPAPPAKPKKEEVPV	6762
QY	649	ETPPPTTSEVSTPTTTTKEPPTTIHKSPTSDBTELSEAPTPKALENSPKRPGVPTTKPAT	708
Db	6763	PVALPQEEVEVPEEEBIVEEEVLPBEEBVEPE--EEBVPPEEEBVEPEEBELPPEEEVPP	6821
QY	709	KPEMTTAKOKTTERDL--RTTPETTTAPAKMTKETATTTEKTESKTIATTTOYSTTT	766
Db	6822	EEBVPPEEEBVEPEEVLPEVKAPKVPVAP-----VPELKKVTVTKKVIPKKEBAPPAK	6876
QY	767	ODTPPKFTTLTKTTTLABRVTTTKTITTTLEIMKPEETAAPKDBATNSKATTPKPOKPT	826
Db	6877	VPEVPKVE--EKRTILPK---EEBVLPEVEETEPEE-----EPISHE	6914
QY	827	KAPKAPTSTKK-----PKTMPRVKR---PKT--TPPRKMTSMPELNTSRIAEMLQTT	877
Db	6915	ELPEEPPSTEEBVEAPPAVPEVIVKAVPEAPTVPKKVEA--PRAKSKATLPEEKVPVP	6972
QY	878	TRPNQTPNSKILEVNPKSEDAGAGETPRHMLRPHVFMPEVTPMDVLPVPPNOGIITN	937
Db	6973	VQKKAPPAKVPEV-----PKVPEKKVLP--VPKKEAV--	7003
QY	938	PMLSETNINCNGPYDGLTTLNGLIYNAR	967
Db	7004	-----PPAKGRTVLEEKVSVAR	7021

RESULT 13  
 C84672  
 hypothetical protein At2g27380 [imported] - Arabidopsis thaliana  
 C.Species: Arabidopsis thaliana (mouse-ear cress)  
 C.Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C.Accession: C84672  
 R.Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
 M.; Koo, H.; Wolfat, K.S.; Cronin, L.A.; Shrop, M.; Vankken, S.E.; Unayam, L.; Tallon,  
 enus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
 Nature 402, 761-766, 1999  
 A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A.Reference number: A84420; MUID:20083487  
 A.Accession: C84672  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-761 <STO>  
 A.Cross-references: GB:AE002093; NID:g5306260; PIDN:AA41892.1; GSPDB:GN00139  
 C.Genetics:  
 A:Gene: At2g27380  
 A:Map position: 2

Query Match	11.1%;	Score 632;	DB 2;	Length 761;
Best Local Similarity	31.1%;	Pred. No. 9.7e-21;		
Matches 217;	Conservative	44;	Mismatches 354;	Indels 82;
			Gaps	33;
QY 132	PRPKAETTKGPAITLP---KEPTLP-TPEKPASTPKPEPTTIKSAPTTKEKPAVTT	187		
Db 69	PPPIQKPPPIYSPPIYPPPIQKPPPIYSPPIYPPPIQKPPPIYSPPIYPPPIQKPPPT	128		
QY 188	KSAPTTPEKEBAPTTKEPAP--TTKEKPAVTTTKEBAPTTTKSAPITLP---KEKAPV--	239		
Db 129	YSPPIYV---PPIQKPPPIYSSPVPKPPPV---QMPPIYSSPIKPPVHNKPPPIYTS	181		
QY 240	TRKKP---APTTKEKPAVTP---KEPVP--TTPEKPAVTTKEBAPV--TPKEBAPVAK	289		
Db 182	PRPKPVHNKPPPIYSPPIKPPVHNKPPPIYSPPIKPPVHNKPPPIYSPPIKPPVHNK	241		



A:Reference number: 151618; MUID:96019267  
A:Accession: 151618  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-990 <CAI>  
A:Cross-references: EMBL:X88927; NID:9895920; PIDN:CAA61368.1; PID:9895921  
C:Genetics:  
A:Gene: xNoppl80  
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein  
C:Keywords: phosphoprotein

Query Match 11.0%; Score 626.5; DB 2; Length 990;  
Best local similarity 26.2%; Pred. No. 2.2e-20;  
Matches 253; Conservative 132; Mismatches 393; Indels 187; Gaps 45;

```
QY 3 DNKKNTTKKKPPPPVVDAGSGLDNGDFKTTPTDTSTQHNKVTSPKITTAKPINR 62
Db 59 DAKRRRPANGLPKKRSKSSSESDSEDEPPAKKRAQ-----PAGKKPVKAVQPK 114
QY 63 PSLPNSDTSKETSLVNNKETTVE--TKETTTNK-----QSTDGKKTTSKETOIEK 116
Db 115 KAKSSSESDSESD-----SEETTKPPAKRPAQTPKVAAVKTPQKKAKSSS-ESSSED 170
QY 117 TSAKDLPATSKVLAKPPKAETTTKGPALTTKEPPTTPEK-----PASTTKEPTPTI 172
Db 171 EAKSKKOPVKKV-----PPKQAVVAGIASNNGKTADSSSESDSPAKKTAATKTPPT- 225
QY 173 KSAPTTPEKAPATTTSAPTTKEPATTKEPATTKEPATT-KEPATT-----T 226
Db 226 --KPAATAKPAOKTAKGKSSSESDSEDEQAKSKPKPDVYSAVPPPTSVMKKKT 283
QY 227 KSAPTTPEKAPATT-----TPKKAPTTPEKAPATT-----KEPTTPEKAPATT 272
Db 284 LSGPGTAKESSSDSSDEEQPAKAKIYPAKAASAPKPLAKKAETSTDSESDSS 343
QY 273 KEPAATTPEKAPATPKKAPATPKKAPATT-----KEPAP-----TT 311
Db 344 EDEKSSSVKLGVKAAPK--APAAP--DAKSTPVAAKKSAAPAKKASSSDSSSNEETT 400
QY 312 TKESPPTPEKAPATTTSAPTTKEPATTTSAPTTKEPSPPTTKEPATTPEKAP 370
Db 401 TKPAKATTTPKASATPTSKPTNTNG---ATPSTKPAKPGTPTSTAKKSSSDSS 457
QY 371 ---PTTPEKAPATPKKAPAT--TPKEPATTTPKAPATKAPKA-PTTPEKTA----- 418
Db 458 SDEETTTKPAKTTPAKSAATPTSKTP---TNSKATPTSKTTPAKPGTPTSAKKDSS 514
QY 419 -----PTTPKTLPTTEPKLAPTTPEKAPATTPEELAPTTPEEPTPTTPEEPAATP-KA 472
Db 515 SDSSSDSSDEKKTVA--KRAKTTPAKA-----AKTTPAKA-----AKTTPAKP 558
QY 473 AAPNTPKEPAPT-----TPKEPATTPEKAPATTPEKETAATP--KGIAPTKEPATTPK 527
Db 559 AAKSTPGQVPTKKSSSDSSSDSSSEDEKKSAPKAVKTPGKATK-----KPVVAS 611
QY 528 KPAPKEAPATTKEPTSTTSKAPATTPKGTAPTTPEKAPATT--PKEPATTTPKGTAPT 585
Db 612 KPVPAK-----KASSSDSDSSSEETTTKTKPLKLSPAVKTLPPKAESSSDSSSDSD 665
QY 586 TLKEPAPTTPKKAPKELAPTTTGP-----TSTTSDKPAPTTPKETAPT--TPKEP 635
Db 666 SEKK---TKPAKPPAKSATPVNTKAPAKONKASKASCSDDSSSEEGKSQPTGKSPAAK 722
QY 636 APTTPEK--PAPTPEPTPTTSEVSTPTTKE--PTTIHKSPEDESTPELSAAPTPEKALENS 693
Db 723 ATAAPKKNPVAVNKDPSSSSSDSSGDEKQPKQAAAARDVKGAKAAKPTPKKAAS 782
QY 694 PKPEGPVTTTAAAKPEMTTAKDKTTERDLRTTETTTAAPKMTKETATTEKTESK 753
Db 783 SSE---DSSSDSDVSKAKKTNIAVSKSPV---TTPKAVPAKAKSSSESDSEDEKOG 835
QY 754 IYATTTQVYSTTQDTTPEFKITTLKTTTLAPKVTITK--KITTTPEIMNKPPEETA-KPKD 810
```

```
Db 836 KNISTTKIANST-----PKAAAAECSESSSSSEDEGKANGTSGKRR 877
QY 811 RAINS---KATTPKQKPTKAPKPKYSTKKPKTTPRVRKKTPTP--RKNTSTIMPLNPT 866
Db 878 ESTGNACEAVTPE-----NKKLKAKSPNTPFPKVNKKELANTPPRRVVEEDIEINP- 928
QY 867 SRIAE 871
Db 929 -RMAD 932
```

Search completed: April 26, 2002, 16:19:22  
Job time: 352 sec

Mon Apr 29 08:35:39 2002

us-09-556-246-1\_copy\_200\_1263.rpr

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:35 ; Search time 47.4 Seconds

(without alignments)  
823.025 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1263

Perfect score: 5698  
Sequence: 1 VKDNKKRRKKRPKPPVY.....VAALSTAKYKMPESYVFEK 1064

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1168	20.5	5179	1 MUC2_HUMAN	002817 homo sapien
2	950	16.7	1664	1 SLPI_CLODM	006852 clostridium
3	789.5	13.9	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	11.4	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	556.5	9.8	875	1 FPI_MYTEO	Q25460 mytilus edu
6	555.5	9.7	2700	1 ZAN_HUMAN	Q9Y453 homo sapien
7	551	9.7	620	1 EXTN_TOBAC	P13963 nicotiana t
8	533	9.4	1087	1 NEH_MOUSE	P13963 nicotiana t
9	530.5	9.3	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	9.3	865	1 CPN_DROME	Q02910 drosophila
11	518.5	9.1	872	1 FPI_MYTEO	Q25434 mytilus cor
12	509.5	8.9	662	1 MUC1_XENLA	Q05049 xenopus lae
13	503.5	8.8	1970	1 RPB1_HUMAN	P24928 homo sapien
14	498.5	8.7	1970	1 RPB1_MOUSE	P08875 mus musculu
15	497	8.7	831	1 NEH_RAT	P16884 rattus norv
16	493.5	8.6	467	1 RPB1_CRIGR	P11414 cricetus
17	488.5	8.6	836	1 SSP2_PLATO	Q01443 plasmodium
18	488	8.6	1020	1 NEH_HUMAN	P12036 homo sapien
19	475.5	8.3	267	1 EXTN_MAIZE	P14918 zea mays (m
20	471.5	8.3	5376	1 ZAN_MOUSE	P08879 mus musculu
21	468.5	8.2	634	1 HMP1_CANAL	P46593 candida alb
22	467	8.2	817	1 VPR1_YEAST	P37370 saccharomyc
23	463	8.1	2142	1 BAT2_HUMAN	P48614 homo sapien
24	454	8.0	797	1 VGLX_HSVB	P28958 equine herp
25	454	8.0	1161	1 YJ9P_YEAST	Q00130 ictalurus gal
26	448.5	7.9	670	1 VGLX_HSVB	P28958 equine herp
27	442.5	7.8	751	1 FPI_MYTEO	Q27409 mytilus gal
28	439.5	7.7	1083	1 XPD2_HUMAN	P17437 xenopus lae
29	432	7.6	439	1 XPD2_XENLA	P10220 xenopus lae
30	432	7.6	3164	1 TEGU_HSV1	Q9um66 homo sapien
31	426	7.5	2715	1 TRX2_HUMAN	P27546 mus musculu
32	424.5	7.4	1125	1 MAPA_MOUSE	P02840 drosophila
33	424	7.4	307	1 SGS3_DROME	

## ALIGNMENTS

RESULT	ID	SEQUENCE	STANDARD	PRT	5179 AA.
1	MUC2_HUMAN	002817, Q14878, 01-JUN-1994 (Rel. 29, Created)			
AC	DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	DE	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	GN	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).			
OS		MUC2 OR SMUC.			
OC		Homo sapiens (Human).			
OC		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC		Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX		NCBI_TaxID=9606;			
RN		[1]			
RP		SEQUENCE FROM N.A.			
RC		TISSUE=Intestine;			
RX		MEDLINE=94132002; PubMed=8300571;			
RA		Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;			
RT		"Molecular cloning of human intestinal mucin (MUC2) cDNA.			
RT		Identification of the amino terminus and overall sequence similarity			
RT		to prepro-von Willebrand factor ";			
RL		J. Biol. Chem. 269:2440-2446(1994).			
RN		[2]			
RP		SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.			
RC		TISSUE=Colon;			
RX		MEDLINE=93016075; PubMed=1400449;			
RA		Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,			
RT		Petersen G.M., Kim Y.S.;			
RT		"MUC-2 human small intestinal mucin gene structure. Repeated arrays			
RT		and polymorphism ";			
RL		J. Clin. Invest. 88:1005-1013(1991).			
CC		- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND			
CC		OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A			
CC		PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS			
CC		AGENTS AT MUCOSAL SURFACES.			
CC		- SUBUNIT: MULTIMERIC.			
CC		- SUBCELLULAR LOCATION: SECRETED.			
CC		- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,			
CC		BROCHUS, CERVIX AND GALL BLADDER.			
CC		- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR			
CC		INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).			
CC		- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND			
CC		VARIABLES AMONG DIFFERENT ALLELES.			
CC		- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT			
CC		OF SILKWORM HEMOCYTIN.			
CC		- SIMILARITY: CONTAINS 2 WFGC DOMAINS.			
CC		- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).			

DR EMBL; L21998; AAB95295.1; -.  
DR EMBL; M74027; AAA59875.1; -.  
DR EMBL; M94131; AAA59163.1; -.  
DR EMBL; M94132; AAA59164.1; -.  
DR MIM; 158370; -.  
DR InterPro; IPR000359; Cys\_knot.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002400; GF\_cys\_knot.  
DR InterPro; IPR001007; VMFC.  
DR InterPro; IPR001846; Vwd.  
DR Pfam; PF00007; Cys\_knot; 1.  
DR Pfam; PF00094; vwd; 4.  
DR PRINTS; PR00438; GFCSKNOT.  
DR SMART; SM00214; VMC; 2.  
DR SMART; SM00011; VMC\_def; 2.  
DR SMART; SM00216; VWD; 4.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.

DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01225; CTCK\_2; 1.  
DR PROSITE, PS01208; VMFC; 2.  
KW Glycoprotein; Repeat; Signal.

FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	5179	MUCIN 2.
FT	DOMAIN	1401	1747	APPROXIMATE REPEATS.

FT	1401	1.
REPEAT	1416	
REPEAT	1417	2.
REPEAT	1432	
REPEAT	1433	3.
FT	1448	
FT	1450	4.
REPEAT	1454	

FT	1470	72
REPEAT	1404	
FT	1472	6.
REPEAT	1465	
FT	1465	5.
REPEAT	1404	4.

FT	145	7B.
REPEAT	1517	
FT	1533	8A.
REPEAT	1534	
REPEAT	1556	8B.

FT	REPEAT	1557	1572	9A.
FT	REPEAT	1573	1596	9B.
FT	REPEAT	1597	1612	10A.

FT	1613	1635	10B.
REPEAT			
FT	1636	1651	11A.
REPEAT			
FT	1652	1675	11B.
REPEAT			

FT	REPEAT	1676	1683	12.
FT	REPEAT	1684	1699	13.
FT	REPEAT	1700	1715	14.

FT	1716	1731	15.
REPEAT	1732	1747	16.
FT	4815	4886	VMFC 1.
DOMAIN			

FT	DOMAIN	4924	4991	VMFC. 2.
FT	DOMAIN	5075	5160	CTCK.
FT	DISUFLID	5075	5122	BY SIMILARITY.

FT	5089	BY SIMILARITY.
DISOLEID	5136	BY SIMILARITY.
FT	5098	BY SIMILARITY.
DISOLEID	5152	BY SIMILARITY.
FT	5102	BY SIMILARITY.
DISOLEID	5154	BY SIMILARITY.

	BY SIMILARITY.	
FT	DISULFID	?
FT	CARBOHYD	163
FT	CARBOHYD	423
FT	CARBOHYD	423
	N-LINKED (GLCNAC . .)	(POTENTIAL) .
	N-LINKED (GLCNAC . .)	(POTENTIAL) .
	N-LINKED (GLCNAC . .)	(POTENTIAL) .
	N-LINKED (GLCNAC . .)	(POTENTIAL) .

FT	6/0	-	(POTENTIAL)
CARBOHYD	7/0	-	(POTENTIAL)
CARBOHYD	7/0	-	(POTENTIAL)
CARBOHYD	8/4	-	(POTENTIAL)
CARBOHYD	8/4	-	(POTENTIAL)
FT	13/26	-	(POTENTIAL)

FT CARBOHYD	1139	.	(POTENTIAL).
FT CARBOHYD	1154	N-LINKED	(POTENTIAL).
FT CARBOHYD	1215	N-LINKED	(POTENTIAL).
FT CARBOHYD	1230	N-LINKED	(POTENTIAL).
FT CARBOHYD	1230	N-LINKED	(POTENTIAL).

[illegible]

Query Match	20.5%	Score 1168;	DB 1;	Length 5179;
Best Local Similarity	32.5%	Pred. NO. 1.9e+40;		
Matches 363;	Conservative 53;	Mismatches 448;	Indels 254;	Gaps 40.

QY	21	DEAGSGLDNGDFK-----	VTTPDTSTTQH-NKVSTSPK-----	52
Dh	1312	DHAGSGDGDGDRPEPDGCGAPEDI	ICRSYKVDLHSLIEHGKQCDVSV	GLTCKNEDEQ
	1371			

53 -----TTAKPINRPSLPNSDTSKETSJLVNKETVEIK 88  
||| ||| : : : : : ||

Db 1372 GNGFGLCYDYKIRIVNCCGPMDDKITTSPSPPTTTPSPPTTTTTLPLPTTTPSPPTT----- 1437

QY 89 ETTTNNKQSTGDEKRTTSAKETOSIEKTSAKDLAPTSTVLAKTTPKAEETTTKGPAITTP 148

```

      1428 -TTTTTTTTTSPPTTTTTT-----LPTT-----TSPBISTTTTTTPTPTTTP 1470

```

QY	149	KEPT	-----	PTTPKEPAS	TPKEKEPT	TKSAP	-----	TPKKE	-----	APLTKNSAF	116	134
Db	1471	SPPTTTPSPPTT	SPPTT	TTTTT	PPPTT	SPPKMT	TTTT	PPAS	TTTT	LLPPTT	SPPTT	1530

[illegible]

Db	1531	TTTPTTTTSSPPIITPLIPPSITLTPPLIIPSPFIILIIIEFTLLICGFLACGG... 2926
QY	246	PTTKKPAAPPTPKKEPTPTTKKPAATTKKBPATPKP-----APTAKKKPATPTPK 298

Db 1591 ITTTTTPTTTTSPPTTTTTTTTPTTPTTSPPTTTITPTTSTTLTLPPTTTTSPPTTTTT 1650

[illegible]

QY	356	TTTKEPAPITPKEPAPITPKKAPAPITPKEPAPITTTKKKAPAPAPKEPAPITP-	414
Dh	1710	TTTTTSSSTTSPSPPTTMTTSPSPPTTSPPTTMTTLPSPPTSSPPTTLPSPSTTPT	1769

QY 415 ---KETAPTP-----KKLTP 427  
| | | | : |

Db	1770	FSFSTTPTTPPCVLCNWTGMDSGKPNRHKPGDTELLGDVCGGMAANISCRATMYP	1829
OY	428	-----TTPKLIAPT	437

Db 1830 DVPIGLGQTVVCDYSVGLICKNEDQKPGVYIPMAFLNFEINVOCCCECVYQPTTMTT 1889

QY	438	PEKPAPTPEELAPTPPEEPPT -PPEEPAPTPKAAAPNPKKEPAPTPPEP -	----	AV	491
		:	:		
Db	1890	TENPPEPTTPTTPTTVPPTPTGQTPTTPTTTTIVTPTPTPTGTOTPTTPTTT			1949





```

OY 857 ---TSMPELNPFSRIAEAMLOTTTRPNPTSPKSLVEVAPKXSDGAGCEGPH- 908
Dd 1454 YLRGPDGSPFRPERNITRADAAVIF-----AKLL-----GADESGAOSAPSYSD 1498
OY 909 LLRPH-----VEMPE-----VTPMDYLPRVPGGT----- 934
Dd 1499 LADTHMAAMAATKFAATSGLRKGYPDGFRKDDONTTRAEFAATVILHFLTKVCGEIMSKLA 1558
OY 935 ---IINPLSDETNICNG---KPYDGLTL-----RNGTLVAFRGHTFMMLSPSPS 981
Dd 1559 TIDISNPFEDD---CVGHMAQAEFIKLTSLGYSIGYDPT-----FKPON 1600
OY 992 PARRTLEWGIPIPIDIVFTTCNCEGKTFPEKD--SOQWFFTDNDIKD 1026
Dd 1601 YIKRSESY---ALINRALERGLNGAKPLFPDYNESTMAF-GDIND 1642

RESULT 3
AMYL_YEAST STANDARD; PRT; 1367 AA.
ID AMYL_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUCOAMYLASE SI/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
  GLUCOSYLASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
DE STAI OR STAZ OR MA15 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SZ288C / AE972;
RA Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churchill C.M., Connor R., Copley T., Dear S., Devlin K., Fraser A.,
RA Gentes S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Raebordneam M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RC MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
  STAI."
RT J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RA MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STAZ and SGA genes
  from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
-1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
  GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
  WITH RELEASE OF BETA-D-GLUCOSE.
CC -1- SIMILARITY: TO S.POMBE SPC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38061; CA86176.1; -
DR EMBL; M16164; AAA35014.1; -

```

Query Match	13.9%	Score 789.5	DB 1	Length 1367
Best Local Similarity	28.4%	Pred. No. 1.2e-25		
Matches 301	Conservative 104	Mismatches 483	Indels 171	Gaps 47
DR EMBL: M16165; AAA35015.1; -				
DR EMBL: X13857; CAA32069.1; -				
DR PIR: B26877; B26877-				
DR PIR: A26877; A26877-				
DR PIR: S48478; S48478-				
DR SGI: S0001458; M0C1-				
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;				
KW Signal; Multigene family.				
FT SIGNAL	1	21	POTENTIAL.	
FT CHAIN	22	1367	GLUCAMYLASE SI/S2.	
FT CDS	210	1367	SER/THR-RICH.	
FT CAROIND	817	817	N-LINKED (GLCNAc. . .) (POTENTIAL).	
FT CAROIND	874	874	N-LINKED (GLCNAc. . .) (POTENTIAL).	
FT CAROIND	1367	1367	AA; 136110 MW; 91C00E2BDB61A9D CRC64;	
SEQUENCE				
Query Match	13.9%	Score 789.5	DB 1	Length 1367
Best Local Similarity	28.4%	Pred. No. 1.2e-25		
Matches 301	Conservative 104	Mismatches 483	Indels 171	Gaps 47
DR EMBL: M16165; AAA35015.1; -				
DR EMBL: X13857; CAA32069.1; -				
DR PIR: B26877; B26877-				
DR PIR: A26877; A26877-				
DR PIR: S48478; S48478-				
DR SGI: S0001458; M0C1-				
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;				
KW Signal; Multigene family.				
FT SIGNAL	1	21	POTENTIAL.	
FT CHAIN	22	1367	GLUCAMYLASE SI/S2.	
FT CDS	210	1367	SER/THR-RICH.	
FT CAROIND	817	817	N-LINKED (GLCNAc. . .) (POTENTIAL).	
FT CAROIND	874	874	N-LINKED (GLCNAc. . .) (POTENTIAL).	
FT CAROIND	1367	1367	AA; 136110 MW; 91C00E2BDB61A9D CRC64;	
SEQUENCE				

DB 967 VPTTTTSSSTTTTTCSTGNSAGETSCGSPKTTTTCSTSPSELA----- 1021  
OY 812 ATNSKATTPKPKP-----TKAPKKPSTTKPKMPRRKKTPTPR 854  
DB 1022 ---SESTTSPPTPVTVTVSTVTVTEYSTPGEITTTTPTVTLTTP- 1077  
OY 855 KMTSTPELNPSTRIAEAMKQTTTRPNQPNKSLVEVNPKSEDAEGAEGTPIMLRPHV 914  
DB 1078 -SVTITVNEPPT-ITTVVCSSTGT-----NSAGETSCGSPKTTTTCST 1122  
OY 915 FMEVPPDDOYLPRVNOGIINPMLSDEFINICNGKPVGCTTLRNGTLVAFRGHYFMKL 974  
DB 1123 GTEETVTEATLTVTAVTTVTSTGTSNSTA-GKTTGYTKSVPT-----VYTL 1175  
OY 975 SPSPSPPARRTVEWGISPIDVFTRC-----NCEGKT 1009  
DB 1176 APSAPVTPATN-----AVPTTIT--TECSAATNAGET 1207

## RESULT 4

MUC1\_HUMAN

ID MUC1\_HUMAN STANDARD; PRT: 1255 AA.

AC P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;

DT 01-JAN-1990 (rel. 13, Created)

DT 01-APR-1990 (rel. 14, Last sequence update)

DT 20-AUG-2001 (rel. 40, Last annotation update)

DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)

DE (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)

DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H2TAG) (PEANIT-

DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN

DE DF3).

GN MUC1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE-Pancreas; PubMed-2394722;

RA MEDLINE-90368716; Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;

RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";

RL J. Biol. Chem. 265:15294-15299(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-90202794; PubMed-2318825;

RA Lichtenberg M.J.T., Vos H.L., Gennissen A.M.C., Hilkens J.;

RT "Epistatin, a carcinoma-associated mucin, is generated by a

RT polymorphic gene encoding splice variants with alternative amino

RT terminl.";

RL J. Biol. Chem. 265:5573-5578(1990).

RN [3]

RP SEQUENCE FROM N.A.

RX TISSUE-Breast carcinoma;

RA MEDLINE-90368715; PubMed-1697589;

RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,

RT Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;

RT "Molecular cloning and expression of human tumor-associated

RT polymorphic epithelial mucin.";

RL J. Biol. Chem. 265:15286-15293(1990).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE-91097524; PubMed-2268309;

RA Lancaster C.A., Peat N., Duhig T., Wilson D.,

RA Taylor-Papadimitriou J., Gendler S.J.;

RT "Structure and expression of the human polymorphic epithelial mucin

RT gene: an expressed VNTR unit.";

RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).

RN [5]

RP SEQUENCE FROM N.A.

RX TISSUE-Breast carcinoma;

RA MEDLINE-90276413; PubMed-2351132;

RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,  
RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,  
RA Keydar I.;

RT "Human epithelial tumor antigen cDNA sequences. Differential splicing  
RT may generate multiple protein forms.";

RL Eur. J. Biochem. 189:463-473(1990).

RN [6]

RP SEQUENCE FROM N.A.

RX TISSUE-Breast carcinoma;

RA MEDLINE-90276414; PubMed-2112460;

RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horov J.,

RA Zilhan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;

RT "A transcribed gene, containing a variable number of tandem repeats,

RT codes for a human epithelial tumor antigen. cDNA cloning, expression

RT of the transcribed gene and over-expression in breast cancer

RT tissue.";

RL Eur. J. Biochem. 189:475-486(1990).

RN [7]

RP SEQUENCE FROM N.A.

RX MEDLINE-91033045; PubMed-1688329;

RA Tsarfaty I., Hareuveni M., Horov J., Zaretsky J., Weiss M.,

RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;

RT "Isolation and characterization of an expressed hypervariable gene

RT coding for a breast-cancer-associated antigen.";

RL Gene 93:313-318(1990).

RN [8]

RP PARTIAL SEQUENCE FROM N.A.

RX MEDLINE-88330762; PubMed-3417335;

RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,

RA Burchell J.;

RT "A highly immunogenic region of a human polymorphic epithelial mucin

RT expressed by carcinomas is made up of tandem repeats.";

RL J. Biol. Chem. 263:12820-12823(1988).

RN [9]

RP SEQUENCE OF 1-169 FROM N.A.

RX MEDLINE-90088473; PubMed-2597151;

RA Abe M., Siddiqui J., Kuft D.;

RT "Sequence analysis of the 5' region of the human DF3 breast

RT carcinoma-associated antigen gene.";

RL Biochem. Biophys. Res. Commun. 165:644-649(1989).

RN [10]

RP SEQUENCE OF 1-109 FROM N.A.

RX TISSUE-Thyroid;

RA MEDLINE-96183746; PubMed-8608966;

RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;

RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse

RT transcriptase polymerase chain reaction of the MUC1 gene.";

RL Int. J. Cancer 66:35-39(1996).

RN [11]

RP SEQUENCE OF 1-89 FROM N.A.

RX TISSUE-Lung;

RA MEDLINE-96181716; PubMed-8604237;

RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,

RA Lee L.N., Lun K.T., Wu C.W.;

RT "Mucin mRNA expression in lung adenocarcinoma cell lines and

RT tissues.";

RL Oncology 53:118-126(1996).

RN [12]

RP SEQUENCE OF 1-46 FROM N.A.

RX TISSUE-Breast carcinoma;

RA Bulwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;

RT Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.

RT "- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN

RT CYTOSKELETON.

RT "- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM

RT IS ALSO PRODUCED.

CC "- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE

CC SPLICING.

CC "- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL

CC TUMORS, SUCH AS BREAST CANCER.

CC "- PTM: HIGHLY GLYCOSYLATED (N- AND O-LINKED CARBOHYDRATES AND SULFIC

CC ACID).

CC "- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT

CC	VARIS FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
CC	MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
CC	-1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC	This SWISS-PORT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb.sib.ch">license@isb.sib.ch</a> ).
CC	-----
DR	EMBL; J05582; AAA60019.1; -
DR	EMBL; M32738; AAA35804.1; -
DR	EMBL; M32739; AAA35806.1; -
DR	EMBL; J05581; AAA59876.1; -
DR	EMBL; M61170; AAB53150.1; -
DR	EMBL; X52229; CAA36478.1; ALT SEQ.
DR	EMBL; X52228; CAA36477.1; ALT SEQ.
DR	EMBL; M35093; AAB59612.1; ALT SEQ.
DR	EMBL; Z117324; CAA78972.1; -
DR	EMBL; Z117325; CAA78973.1; -
DR	EMBL; M31823; AAA35757.1; -
DR	EMBL; S81781; AAD14376.1; ALT INIT.
DR	EMBL; S81736; AAD14369.1; ALT INIT.
DR	EMBL; M21868; AAB59874.1; ALT SEQ.
DR	PIR; A35175; A35175.
DR	PIR; B35175; B35175.
DR	PIR; S10218; S10218.
DR	GLCOSuitedb; P15941; -
DR	MIM; 158340; -
DR	MIM; 113720; -
DR	InterPro; IPR000082; SEA.
DR	Pfam; PF01390; SEA. 1.
DR	SMART; SM00200; SEA. 1.
DR	PROSITE; PS50024; SEA. 1.
KW	Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW	Repeat; Alternative splicing.
FT	SIGNAL 1 23
FT	POTENTIAL.
FT	MUCIN 1.
FT	CHAIN 24 1255
FT	DOMAIN 24 1162
FT	TRANSMEM 1163 1186
FT	DOMAIN 1187 1255
FT	DOMAIN 81 960
FT	DOMAIN 1034 1151
FT	CARBOHYD 957 957
FT	CARBOHYD 975 975
FT	CARBOHYD 1029 1029
FT	CARBOHYD 1055 1055
FT	CARBOHYD 1133 1133
FT	VARSPPLIC 19 19
FT	VARSPPLIC 20 22
FT	VARSPPLIC 20 31
FT	VARSPPLIC 126 905
FT	VARSPPLIC 1077 1087
FT	VARSPPLIC 1088 1255
FT	CONFLICT 2 2
FT	CONFLICT 134 134
FT	CONFLICT 134 134
FT	CONFLICT 1021 1021
FT	CONFLICT 1251 1251
FT	CONFLICT 1251 1251
FT	SEQUENCE 1255 AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;
SO	

Query Match 11.4% Score 651; DB 1; Length 1255;

Best Local Similarity 27.8%; Pred. No 4.4e-20;

Mismatches 299; Conservative 72; Mismatches 460; Indels 234; Gaps 51

Score 651; DB 1; Length 1255;

Pred. No 4.4e-20;

Mismatches 460; Indels 234; Gaps 51

```

Query Match 11.4% Score 651 DB 1 Length 1255;
Best Local Similarity 27.8% Pred. No. 4.4e-20;
Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps

QY 75 TSLVYNKETTVEKETTNNKOTSDCEKETTSAKETDISIKTISAKD-LAPTSVYLAKPT 133
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 16 TLYLV-----VTSQGHASSPPGGCKETSATQRSVSSSTPEKNAVSMITSSVLSHS 65
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

QY	134	P-BAEYTTKGP--ALTTKPEP-----	TPPTKPEBASTTKKEPTPTT---	K	173
Db	66	PGSGSSSTPGODVTLAPATEPAPSGSAATWGODVTSVPTPRALGSTITPRADHVSADPNK		125	
QY	174	SAPPTKPEPAPTTKSAPPTKPEAPTTTKEP-----	APTTKPEAPTTTKKEPAPTTTK	227	
Db	126	PAPGSTAPRAHGVY- SAPDT--	RAHGVSTAPRAHGVTSAPDT--	RAHGVSTAPRAHGVY- 179	
QY	228	SAPPTKPEAPTTT-----	KKRABTTKPEP-----	APTTKKEPTPTT-----	265
Db	180	SAPDTRAPGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVTSAPDTRAPAGSTAPRAHGV		239	
QY	266	---KEAPPTKPEP-----	APTTKKEAPAPAK-----	KRAPTTKPEAPTTTK-----E	306
Db	240	SAPDTRAPGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVTSAPDTRAPAGSTAPRAHGV		299	
QY	307	PAPPTKPEPPTTKPEAPTT---TKSAPTTKPEAPTTKSAPPTKPEPPTTKKEP--		361	
Db	300	SAPDTRAPGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVY- SAPDT--	RAHGVSTAPRAH	356	
QY	362	---APTTKPEAPTTTK-----	KRAPTTKKEAPAPTTK-----	EPATTTKKRAPAPAKPEP	409
Db	357	GVTSAPODTRAPAGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVTSAPDTRAPAGSTAPRAH		416	
QY	410	APPTKPEAPTTKPKLLPTTPEKLAPTTPEKAPPTAPEELAPTTPEEPPTTPEEPAPT		469	
Db	417	GVTSAPODTRAPAGSTAPRAHGVTSAPDTRAPAGSTAP--	RAHGVTSAPDTRAPAGSTA	472	
QY	470	PKA-----AAPTKPEAPAPTTK-----	EPADTTKPEAPPTTKETAAPTTKGAAPTTLKPEP	521	
Db	473	PRAHGVTSAPDTRAPAGSTAPRAHGVTSAPDTRAPAGSTAP--	RAHGVTSAPDTRAPAP	528	
QY	522	APTTP-----	KKRAPKELAPTTKEPTSTSOKRAP--	TTKGAAPTTKPEAPPT	569
Db	529	GSTAPRAHGVTSAPDTRAPAGSTAP--	RAHGVTSAPDTRAPAGSTAPRAHGVYS----	ARD	583
QY	570	TKPEAPPTPKG---TAPTTLKEAPAPTT-----	KKRAPKELAPTTKGPTSTT	615	
Db	584	TPRAGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVTSAPDTRAPAGSTAP--	RAHGVTSAP	642	
QY	616	SOKRAP--TTK-----	ETAPTTKPEAPPT-----	KKRAPPTTPTPEPTTSEVS	659
Db	643	DTRAPAGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVTSAPDTRAPAGS--	TAPRAHGVYS	700	
QY	660	PTTTTKKEPTT-----	HKSPDSEPTBELSEPTKALENSPEKGVPTTKTPAATKPEM	712	
Db	701	APDTRAPAGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVTSAP--	DTRAPAGSTAP--	753	
QY	713	TTTAAKDTTERDLRTTPEPT-----	TAADKMTKETATTEKTESKITATTTYTSTT	765	
Db	754	--RAHGVTSAPDTRAPAGSTAPRAHGVTSAP--	DTRAPAGSTAPRAHGVTSAPDTRAPRA	807	
QY	766	TODDTP--FKITTLTKTTLAKVTTTKTITTT--	ELMNKPEETAKKDKATSKATPK	821	
Db	808	PGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVTSAPDTRAPAGSTAPRAHGVTSAPDTRAPRA		867	
QY	822	PKKPTAPKPKPTSTKKPKTMEPVKKPKTPTPKMTSTMEPLNSTRSIAEAMLOTTTRPN		881	
Db	868	P--GSTAPRAHGVTSAPDTRP--	APGSTAPRAHGVTSAPDTRARA-----	PG	909
QY	882	QT--PNSKLEVNPKSEDAAGABETTRHMLLRPHVFMPEVPTTDMOYLRLVRPQGLINPML		940	
Db	910	STAPRAHGVTSAPDTRAPAGSTAPRAH--	GVTSAPODTRAPALGSTA--	PPV	955
QY	941	SDETNITCNCKPVDGLTTLKNGTLVAFRCHYFMMLSPSPS		981	
Db	956	HNVTSAAGSAGSASTLVHNGTSABATTTPAKSKTPEPSIPS		996	

RESULT	5
FP1_MYTED	
ID	FP1_MYTED
STANDARD;	
PRT;	875 AA

AC 025460.  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADHESIVE PLACOE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT  
 DE PROTEIN 1) (MEFPI) (FRAGMENT).  
 GN FPI.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidae;  
 OC Mytiloidae; Mytilidae; Mytilus.  
 OX NCBI\_Taxid=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=91023829; PubMed=1367451;  
 RA Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;  
 RT "Structural and functional repetition in a marine mussel adhesive  
 RT protein.";  
 RL Biotechnol. Prog. 6:171-177(1990).  
 RL [2]  
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.  
 RX MEDLINE=83135732; PubMed=6298211;  
 RA Waite J.H.;  
 RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and  
 RT hydroxyproline-containing decapeptide in the adhesive protein of the  
 RT mussel, Mytilus edulis L.";  
 RL J. Biol. Chem. 258:2911-2915(1983).  
 CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
 CC SUBCELLULAR LOCATIONS CORE COATED WITH ADHESIVE PROTEINS.  
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
 CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
 CC -1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-T-Y-K IS POST-TRANSLATIONALLY  
 CC MODIFIED AS FOLLOWS: THE SIXTH AND SEVENTH RESIDUES ARE  
 CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYHENTHIALANINE  
 CC (DOPA) DERIVED FROM TYROSINE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X54422; CAA38294.1; .  
 DR InterPro: IPR002964; Adhesive\_plaq.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR PRINTS: PRO1216; ADHESIVER.  
 DR PRINTS: PRO1217; PRICHEXTENS.  
 KM Repeat: Hydroxylation.  
 FT NON\_TER 1  
 FT DOMAIN 67 870  
 FT TANDEN REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-  
 FT P-P-[ST].  
 SQ SEQUENCE 875 AA; 100412 MW; 6EA85312748CACE CRC64;

Query Match 9.88; Score 556.5; DB 1; Length 875;  
 Best Local Similarity 28.7%; Pred. No. 2.1e-16;

Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;

QY 42 TONKVVSTPKITAKINP-----RP--SLPNSDTSKETSIVNKEETVET 87  
 DB 1 TKEHPVKKPKTISADYKPPYQPLKKVVDYRPTKSYPTYG-SKTNVYLPKKKLSYK 59  
 QY 88 KETTTNKGSTDSG--KEKTT--SAKETOSIKTSKADLAFTSKVLAKTTPKAEETTKG 142  
 DB 60 PIKTTYAKTNVPPVYKPKMTYPPYKPKPSYPTTKSKRTYPKKITIYPTTKAKSY-- 117  
 QY 143 PAITTPKEPTTPKE-----PASTTPKEPTPTTKSAPTTP-----KEPATPTTKSAPT 193  
 DB 118 PSSYKPKKTPYKPLKPLVPPYKPKPSYPTTKPKPSYPTTKKTYPSSYKAKKPSY 177

QY 194 KEKAPPTTKKEPA--PTTPKEPA-----PTTTKEPA--PTTKSAPTTPK--EPAPT 240  
 DB 178 P-----PTYKAKSPYPTTKAKSPYPTTKAKPTTKAKPTTKAKSPYPTTKAKPTTY 233  
 QY 241 PKKPA--PTTPKEPA--PTTPKEPT--PTTPKEP--APTKEPAPTTPKEPA--PTAPK 289  
 DB 234 KAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKA 293  
 QY 290 KPA--PTTPKEPA--PTTPKEPA--PTTKEPS--PTTPKEPA--PTTKSAPTTPKEPA-- 339  
 DB 294 KPSYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKPTTY 353  
 QY 340 PTTSAPTTPKESPTTKKEPAPTTPKEPAPTTPKPA--PTTPKEPA--PTTPKEPAP 395  
 DB 354 PSTYKAKSPY--PTTKAKSPYPTTKAKPTTKAKSPYPTTKAKSPYPTTKAKSPY 409  
 QY 396 TTTKPAPTAKKEPAPTTPKEPAPTTPKLTPTTEPAPTTPKEPAPTTPKEPAPTTPPE 455  
 DB 410 PPTTKAKPTTKAKP-----TTPST-YKAPSTP--PSYKAKSPYPTTKAKPTTKA 457  
 QY 456 EPTTPKEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKE--PAPTTPKEPAPTT-- 509  
 DB 458 KPT-----YPTTKA--KPSYPAKYKAKSPYPTTKAKSPYPTTKAKSPYPTTKA 506  
 QY 510 PKGAPTTPKKEPAPTTPKPKAPKELAPTTTKEPTSTSDKPAPTTPKGTAPT--PKE 565  
 DB 507 PKLTKPTTK--PKSPSYPSYKPKTTPPTTK-----PKISYPTTKAKPSY 551  
 QY 566 PAPTTPKEPAPTTPKGT--APTTPKEPA--PTTPKPA--PKELAPTTTKG--PTSTSD 617  
 DB 552 PATYKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTY 611  
 QY 618 KPAPTTPKEPAPTTPKKEPA--PTTPKPA--PTT--PETTPPTTSYVSPPTTKKE-- 666  
 DB 612 KAKSPY--PTTKAKSPYPTTKAKPTTKAKPTTKAKSPYPTTKAKSPYPTTKAKPSY 667  
 QY 667 PTTTHKSPDESTPELSAEPPTKALENSPK-----EPGPTT--KTPA-----ATPEMTT 715  
 DB 668 PPT-YKAKSPYPTTKAKSPYPTTKAKPTTKAKSPYPTTKAKSPYPTTKAKSPYPT 726  
 QY 716 AKDPTTERTDITPETTTAAPTTPKMTKETATTTTEKTESKITATTTQVSTTQDPTPKIT 775  
 DB 727 YKAPTTPKAPTYPTSYKAP--TYKAKPTTPPTTKA-----PSYPTTKAKPSYPT 777  
 QY 776 TLTNTTAPVTTTKITTTTTEIMNKPEEYAKPKDRATNSKATTPKQKPTKAPK--P 832  
 DB 778 TYKSKSYPTTKAKSPYPTTKAPPT--YKPLTYPTTK-----PKSPSYPTTKPTTY 826  
 QY 833 TSTKPKPTMPVRKPKPTTP-----RKMTSMPELNPTR 868  
 DB 827 STYKLRKSPYPTTKSKTSTPTTKKISTYSSYKAKTSTYPAVKPTNR 874

RESULT 6  
 ZAN\_HUMAN  
 ID ZAN\_HUMAN STANDARD; PRT; 2700 AA.  
 AC 09Y493; 000218;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ZONADHESIN (FRAGMENT).  
 GN ZAN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE OF 1-2379 FROM N.A.  
 RX MEDLINE=99018118; PubMed=9799793;  
 RA Glokler G., Scherer S., Schatkevooy R., Boright A., Weber J.,  
 RA Tsui L.C., Rosenthal A.;  
 RT "Large-scale sequencing of two regions in human chromosome 7q22;

RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci  
 RT reveals 17 genes.  
 RL Genome Res. 8:1060-1073(1998).  
 RN [2] SEQUENCE OF 2338-2700 FROM N.A.  
 RP TISSUE-Testis;  
 RC MEDLINE=97271566; PubMed=9126492;  
 RA Gao Z., Haruni T., Gaiders D.L.;  
 RT "Chromosome localization of the mouse zonadhesin gene and the human  
 zonadhesin gene (ZAN)."  
 RL Genomics 41:119-122(1997).  
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA  
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR  
 CC SIGNALING.  
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE  
 CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.  
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE  
 CC ZONA PELLUCIDA.  
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,  
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF  
 CC SPERMATOZOA OR PROMOTING ADHESION TO THE CERVICAL ISTHMUS.  
 CC -1- DOMAIN: THE WMP DOMAINS 2 AND 3 MAY MEDIATE COVALENT  
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).  
 CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 4.5 WMP DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF053356; AAC78790.1;  
 CC EMBL: U83191; AAC51208.1;  
 DR MIM: 602372;  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000998; MAM.  
 DR InterPro: IPR002965; P\_fich\_extensn.  
 DR InterPro: IPR002919; TIL.  
 DR InterPro: IPR003328; TILA.  
 DR InterPro: IPR001007; WWC.  
 DR InterPro: IPR001846; Wvd.  
 DR Pfam: PF00629; MAM; 3.  
 DR Pfam: PF01826; TIL; 4.  
 DR Pfam: PF02345; TIL; 4.  
 DR Pfam: PF00094; wvd; 4.  
 DR PRINTS: PR01217; PRICEXTENS.  
 DR SMART: SM00137; MAM; 2.  
 DR SMART: SM00214; WWC; 1.  
 DR SMART: SM00011; WWC\_def; 3.  
 DR SMART: SM00216; WVD; 3.  
 DR PROSITE: PS01186; EGF\_2; 3.  
 DR PROSITE: PS00740; MAM\_1; 1.  
 DR PROSITE: PS50060; MAM\_2; 4.  
 DR PROSITE: PS50060; MAM\_2; 4.  
 KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.  
 FT NON\_TER 1  
 FT DOMAIN <1 109 MAM 1.  
 FT DOMAIN 112 136 MAM 2 (PARTIAL).  
 FT DOMAIN 161 326 MAM 3.  
 FT DOMAIN 322 446 MAM 4.  
 FT DOMAIN 483 951 MAM 4.  
 FT DOMAIN 953 1065 MAM 5.  
 FT DOMAIN 1066 1454 WVD 1 (PARTIAL).  
 FT DOMAIN 1455 1861 WVD 2.  
 FT DOMAIN 1862 2292 WVD 3.  
 FT DOMAIN 2293 2684 WVD 4.  
 FT DOMAIN ? 74 WVD 5.  
 FT DOMAIN ? 74 EGF-LIKE.  
 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 74 74

FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1618 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1832 1832 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1878 1878 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2136 2136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2505 2505 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2374 2379 NOKMA -> RACPGP (IN REF. 1).  
 FT NON\_TER 2700  
 FT SEQUENCE 2700 AA: 293013 MW: 80E60C0B12277B1 CRC64;  
 SO  
 Query Match 9.7%; Score 555.5; DB 1; Length 2700;  
 Best Local Similarity 32.6%; Pred. No. 5,9e-16;  
 Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;  
 QY 58 PINRPSLPPNSDTSKETSILVNRKTYETVETKTTTNKQTSIDGKERTYSAKETQSIIEKT 117  
 DB 445 PVKVLPELPVSPVSS-----TGSEETGLTENPTISTK-----KPTVSIIEKP 487  
 QY 118 SAKDLAPTSKVLAPPTKAEITTTGPRALTTTKEPTPTTKPASTTKPEPT-----PTTI 172  
 DB 488 SVTTEKPT-----VPKEKPTIPTKEPTISTEK-----PTIPSEKPMPSKPTIPESEKPTIL 539  
 QY 173 KSAPTPKEPAPTTTKSAPTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKSAPT 231  
 DB 540 TEKPTIPEE-KPTIPESEKPTISTEKPTIPEE--PTIPEETTTTMEEPVAPTEKPSIPT 596  
 QY 232 TPKEPAPTPPKPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTKAPATPKRP 291  
 DB 597 --EKPSIPEK---PTISMEETIISTEKPTICEKPTIPEK---PTIPEKTSISPEK- 647  
 QY 292 APTTPKEPAPTPPKAPATTTKEPSTTPPKPAPTTKSAPTTKAPATTTKSAPTPK 351  
 DB 648 -PTTPEE-KPTIPEKPTISTEKPTIPEK-PTISPEKLTIPTEKLTIP--EKPTIPT 701  
 QY 352 EPSPTTPKEPAPTTTPKKEPAPTTPKKPA-----PTTPKEPAPTTTPKKEPAPTTKAPAPAP 406  
 DB 702 EKPTISTEE-PTTPEETTTTISTEKPSIPMEKPTLPTEETTSVABETIISTEKLTIPW-- 757  
 QY 407 KEPAATTPKEKAPT-----TEKKLTIPTEPKA-----PTTPKEKAPTTPEELAPTPTEE 456  
 DB 758 EKPTISTEKPTIPEKPTISPEKLTIP-EKLTIPTEKPTIPEETTTISTEKL--TITPE 814  
 QY 457 PPTTPPEEPAPTTPKAAAPNTPKKEPAPTTPKKEPAPTTPKKEPAPTTKGAAPT 516  
 DB 815 KPTISPEKPTISTEK-----PTIPE-KPTIPE-----ETTISTEKLTIPT 855  
 QY 517 TLKEPAPTTPKKPAKELAPTTKEPTSTT-----SDKPAPTTKGAPTTPKKPAPTT 570  
 DB 856 --EKPTISPEKLTIPTEKPTISTEKPTIPEKLTIPTEKPTIPEKPTIPT--EKLTAIR 911  
 QY 571 PKEPAPTTKGAAPTTKAPATTPKPKAPKELAPTTTGGPSTTSDDKPAPTTPKET--A 628  
 DB 912 PPHPSPTA-TGDAALVMSPHASTPWTSV---ILGTTTSSKSTGMSCP-PHARESCAC 966  
 QY 629 PTTTKEPAPT 638  
 DB 967 PASCKSPRS 976  
 RESULT 7  
 EXTN TOBAC STANDARD; PRT; 620 AA.  
 AC P13983;  
 DT 01-JAN-1990 (rel. 13, Created)  
 DT 01-JAN-1990 (rel. 13, Last sequence update)  
 DT 01-MAR-1992 (rel. 21, Last annotation update)  
 DE EXTENSIN PRECURSOR (CELL WALL, HYDROXYPROLINE-RICH GLYCOPROTEIN).  
 GN HRCGPT3.  
 OS Nicotiana tabacum (Common tobacco).

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 CC NCBI\_TaxId=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, XANTHI; TISSUE=Leaf;  
 RX MEDLINE=90128263; PubMed=2612909;  
 RA Keller B., Lamb C.J.;  
 RT "Specific expression of a novel cell wall hydroxyproline-rich  
 RT glycoprotein gene in lateral root initiation.";  
 RL Genes Dev. 3:1639-1646(1989).  
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN  
 CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE  
 CC MAIN ROOT.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.  
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE  
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN  
 CC GLYCOSYLATED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X13865; CAA32090.1; -  
 CC PIR: S06733; S06733  
 CC DR Repeat; Cell wall; Glycoprotein; Signal; Structural protein;  
 CC KM Hydroxylation.  
 CC FT SIGNAL 1 2  
 CC FT CHAIN 1 620  
 CC FT REPEAT 70 73 EXTENSIN.  
 CC FT REPEAT 148 151 H-A-P-P.  
 CC FT DOMAIN 229 242 H-A-P-P.  
 CC FT REPEAT 229 235 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.  
 CC FT REPEAT 236 242 1.  
 CC FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.  
 CC FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.  
 CC SEQUENCE 620 AA: 65406 MW: 641DD2278AB28524 CRC64;

Query Match 9.7%; Score 551; DB 1; Length 620;  
 Best Local Similarity 27.2%; Pred. No. 2.7e-16;  
 Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

QY 136 AETTK-SPALTP--KEPTPT-----PKPASTPKK--PTPTTSAPPTKREPAP 184  
 DB 24 AETTTGGTGLPPPVTSQPPSSIGLSPSADPTTPPSGHNVPSP---RHAPRRHAYPPP 80  
 QY 185 TTTKSAPTTKPEP-----APTTTKREPAP--TPKREPATTTKEPAPTTKSAPTPE 235  
 DB 81 SHGLHLPSSVGGPRPHNGHLRPSRGFNRPSPVISPNSHPPSGAGRRPSGPHLDSHGCR 140  
 QY 236 PAPTPPKKAPKAP--TPKREPATTTKEPTPTTKKEDAPTTK--EPAPTTKREPAPTK 290  
 DB 141 P-----PSPSHGNAPSGHTRPGOHPRSHRRSPSRGHNHPPRYAQRPTPIYS 193  
 QY 291 PAPTPKREPAPTTKREPAPTTKESPTTK--EPAPTTKSAPTTKKREPAPTTKS-- 345  
 DB 194 PEPV--QPPPT--YSPRPPTHVQPTSPSPSGHOPORPTTHNAPPTTHNAPPTHQPSPL 249  
 QY 346 ---APTTKREPSPPTTKREPAPTTKREPAPTTKREPAPTT--TPKREPAPTTK 399  
 DB 230 RHLPSRPRKOPRPPTYSPPRAYAQSPPPTYSPPPPPTSPSPSPSPSPSPSPSP 309  
 QY 400 KRAP--APKREPAPTTKETAPTKPL--PTT-----PEKLAPTTPEKAPTTPEEL 449  
 DB 310 PPTPTPTSPSPSPSPSPSPPTYSPPPTYLPLPSSPTYSPPPVVYSPPPSPSPPTYL 369  
 QY 450 APPTPEEPPTTPEEPAPTTKKAAP--NTPKREPAPTTKREPAPTTKREPAPTTKETA 506

RESULT 8  
 ID NFH\_MOUSE STANDARD; PRT: 1087 AA.  
 AC P19246; 061959;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)  
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).  
 GN NEFH OR NFH.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89121513; PubMed=3220257;  
 RA Julien J.-P., Cole F., Beaudet L., Sidky M., Flavell D., Grosfeld F.,  
 RA Mushynski W.;  
 RT "Sequence and structure of the mouse gene coding for the largest  
 RT neurofilament subunit.";  
 RT Gene 68:307-314(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89089138; PubMed=3145094;  
 RA Sheindman P.S., Garden M.J., Lees J.F., Lazzarini R.A.;  
 RT "The structure of the largest murine neurofilament protein (NF-H) as  
 RT revealed by cDNA and genomic sequences.";  
 RL Brain Res. 464:217-231(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;  
 RA Garden M.J.;  
 RX Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
 CC -1- PRT: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETIDE K-S-P, NFH IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534  
 CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -





CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT  
CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA  
CC PROTEIN.  
CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.  
CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M61732; AAA30255.1; -  
CC PIR: JH0557; JH0557.  
CC HSSP: P29766; JDL1.  
CC InterPro: IPR002860; BNR.  
CC Pfam: PF02012; BNR. 2.  
CC Hydrolase: Glycosidase; Repeat; GPI-anchor;  
CC Phosphorylation.  
CC DOMAIN 1 457 CYS-RICH.  
CC REPEAT 23 BNR 1.  
CC REPEAT 163 174 BNR 2.  
CC REPEAT 209 220 BNR 3.  
CC DOMAIN 589 588 FIBRONECTIN TYPE-III.  
CC CARBOHYD 342 342 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.  
CC CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 9.3%; Score 530.5; DB 1; Length 1162;  
Best Local Similarity 29.4%; Pred. No. 3e-15;  
Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;

QY 152 TPTTPKE-PASTTPKEPTPTTISAPRTTKEPAATTTKSAPTTPKEPAATTP 210  
DB 599 TPTTPDSSAHSTPTSPVDS--SAHSTPTPADSSAHGTPSPVDSAHGT--PSP 651  
QY 211 KE-PAATTPKEPAATTTKSAPTTPKE-PAATTPKEPAATTPKEPAATTP 257  
DB 652 ADSAAGTPTSPVDSAHSTPTSPVDSAHSTPTSPVDSAHGAPSTPADSSAHGTP 711  
QY 258 KEPTP-TTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPS 316  
DB 712 VDSAHGTPSTPADSSAHSTPTPADSS--AHSTPTPADSSAHSTPTPADSSA 764  
QY 317 PTPKEPAATTPKSAPTTPKEPAATTPKSAPTTPKEPSPTTPKEPAATTPKEPA 372  
DB 765 HGTPSTPADSSAHSTPT--PADSSAHGTPSPVDSAHST--PSPVDSAHGTPS 817  
QY 373 TP-KKPAATTPKEPAATTPKKAAPTAPEKPAATTPKEPAATTPKKAAPTA 427  
DB 818 TPVDSAHSTPTSPVDSAHGTPSPVDSAHSTPTPADSSAHST-PSTPADSSAHGTP 876  
QY 428 TTP-EKLAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 477  
DB 877 STPVDSAHSTPTPADSSAHST-TESTPVDSAHSTPTPADSSAHGTPSTPVDSAHGT 935  
QY 478 PKPAATTPKEPAATTPKEPAATTPKKAAPTAATTPKKAAPTAATTPKKAAP 537  
DB 936 PSTPADSS--AHSTPTPADSSAHST-PSTPADSS--AHSTPTSPVDSAHST 983  
QY 538 TTKPTSTSDKPAATTPKGAATTPKE-PAATTPKEPAATTPKGAATTPKKAAP 596  
DB 984 ----PSTPADSSAHST--PSTPADSSAHSTPTPADSSAHSTPTSPVDSAHSTPS 1033  
QY 597 KPAPELAPTTKPTSTSDKPAATTPK----ETAPPTTPKEPAATTPKKAAPTP 652  
DB 1034 TP-----ADSSAHGTPSTPADSSAHSTPTSPVDSAHSTPTSPADSS--AHGTPSTPA 1084

QY 653 PTTSEVSTPTTPKEPTTINKSPDESTP-ELSAPEPTKALENSKEPEGVPTTPATKPE 711  
DB 1085 DSSAH-STPTPAD-SSAHGTP--STPADSSAHSTP-----STPADSSAN 1125  
QY 712 MT 713  
DB 1126 GT 1127  
RESULT 10  
CPN\_DROME  
ID CPN\_DROME STANDARD: PRT; 865 AA.  
AC 002910;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE CALPHOTIN.  
GN CPN OR CAP.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CANTON-S;  
RA MEDLINE=93165729; Pubmed=8094559;  
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;  
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.",  
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CANTON-S;  
RA MEDLINE=93165730; Pubmed=8434015;  
RA Ballinger D.G., Xue N., Harshman K.D.;  
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds  
RT calcium and contains a leucine zipper".  
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).  
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO  
CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL  
CC OF CA+2 PER MOL OF PROTEIN.  
CC -1- SUBUNIT: HOMODIMER (PROBABLE).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.  
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF  
CC COMPOUND EYES AND OCELLI.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL  
CC DEVELOPMENT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L02111; AAA28405.1; -  
CC PIR: L05080; AAA28420.1; -  
CC PIR: A47282; A47282.  
CC FlyBase: FBgn0010218; Cpn.  
KW Calcium-binding.  
DR EMBL: L02111; AAA28405.1; -  
DR PIR: L05080; AAA28420.1; -  
DR PIR: A47282; A47282.  
KW Calcium-binding.  
FT 36  
FT 43  
FT 64  
FT 76  
FT 100  
FT 126  
FT 154  
FT 160  
FT 534  
FT 699  
FT 703  
FT 721  
A -> AVAPAVVA (IN REF. 2).  
I -> T (IN REF. 2).  
T -> V (IN REF. 2).  
I -> A (IN REF. 2).  
P -> P (IN REF. 2).  
VQ -> AP (IN REF. 2).  
I -> V (IN REF. 2).  
S -> T (IN REF. 2).  
A -> E (IN REF. 2).  
I -> T (IN REF. 2).  
V -> L (IN REF. 2).  
D -> E (IN REF. 2).

RESULT	11	
FP1_MYTCO		
ID	FP1_MYTCO	STANDARD;
AC	025434;	PRT;
DT	30-MAY-2000 (Rel. 39, Created)	872 AA.

```

Oy      19 VDEAGSGLDNDEKVTPTDST---TOL-----NVSUSPKTYVKKPINPPSLPPNS 69
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      29 VGSATISGASACAK-TLPGSHPGSKRVHPYKMMNIPt-pyl-SKSGYPAPYKPGY 84

Oy      70 DTSKETSLTVNKEVTEVKETTTNKOtSDGKEKTSIAKETOtSIAKDLAPtSKYL 129
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      85 YPKRPQOPVYSGKTYNPIYKIAKKLSYAKITTYPAVYAKISYSPSK-----HKIT 139

Oy      130 AKPTPKAEVTTTGGPALTTPKEPT--PTTPEPASTTPKBPPTTIKSAPTPKPEAPATTk 188
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      140 YPTPYK-----PKITY- -PYKOKSYPSPSKPYTYPPYK-----PKITYPPYK 185

Oy      189 SAAT--PKPEAPPT--TTKEPAPTTPEKAPATTPEKAPATTTSAPTTPKPEAPATT--PKK 243
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      186 RKKSYPYKPKATYPTTYPKITYP-----PYKKRPST-----PYKPTTYPPYKPKI 246

Oy      244 PAPTTPKEAPATT-----PKETPTT- -PKEPAPTTKEAPTTPEKAPPT--APKKPAPTTp 286
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



FT DISULFID 573 599 BY SIMILARITY.  
 FT DISULFID 583 598 BY SIMILARITY.  
 FT DISULFID 593 610 BY SIMILARITY.  
 FT DISULFID 621 647 BY SIMILARITY.  
 FT DISULFID 631 646 BY SIMILARITY.  
 FT DISULFID 641 658 BY SIMILARITY.  
 FT DISULFID 641 658 BY SIMILARITY.  
 FT VARIANT 276 276 K -> E.  
 FT VARIANT 354 354 C -> R.  
 FT VARIANT 415 415 T -> A.  
 SQ SEQUENCE 662 AA: 67774 MW: F085277F1ED2ED40 CRC64:

Query Match 8.9%; Score 509.5; DB 1; Length 662;  
 Best Local Similarity 28.1%; Pred. No. 1.4e-14;  
 Matches 223; Conservative 48; Mismatches 276; Indels 247; Gaps 30;

QY 92 TTNKOTSTGCKEKTSAKKEQSIKESIASDIAFTSVLAKPPPKATTTTGGALTTTPKRP 151  
 DB 3 TTTAAVAATGCKDTTAAEGSAAEKTA---AGEVASPT--AAVAATGEDATT----- 51  
 QY 152 TPTTPKEPASTTTPKEPTTIKSAPTTPKEPAPTTSKAPTTPKEPAPTTPKEPAPTTPK 211  
 DB 52 -----AATAAETAGTAAGAPPTTTAPATTAGKAPTTAAATAPTTAAGAPTAT 103  
 QY 212 EPAPTTPKEPAPTTPKS-APTTPKEPAPTTPKAPATTTPKEPAPTTPKEPTTPKEPA- 269  
 DB 104 GKAPATTAAPVPTTAASKAPTAAATHTAAATAAPTAAASAASKERSTSSSEEHCH 163  
 QY 270 --PTTPKEAPT--TPKEPAPTAPKAPATTTPK-----EPAPTTPKEPAPTTPKE 314  
 DB 164 VKSKRMCGSKGTTKKO-----CKNNCCDPKHGSHCFHKKRPGHSHHEHTTTTK- 218  
 QY 315 PSPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 374  
 DB 219 -----APTIOIATTTT--TPTTT-----TTTKATPTT----- 245  
 QY 375 KKPAPTTPKEPAPTTPKEPAPTTPKAPAPAPKAPAPTTPKEPAPTTPKATPTTPKEKLA 434  
 DB 246 -----TTTKAPTPTT-----TTTKATPTT-----TPTT----- 270  
 QY 435 PTPPEKAPTTPPELAPTTPPEPTTPPEBAPTTPKRAAAPTTPKEPAPTTP 494  
 DB 271 -----TTTKATPTTPTTTTPTT----- 289  
 QY 495 KEPAPTTPKEAPTTPKGTATPTLKEPAPTTPKAPKAPKELAPTTTKEPTSTSDKAPAPT 554  
 DB 290 -----TTTTKATTTTTTTTSGECKMEPSK-----REDCGYSGTTESQCR 328  
 QY 555 PKG-----TAPPT-----PKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKELA 604  
 DB 329 TKGCCFDDSIQTKWCYTTLSQVADCKVEPSQVRDCCGRGIT---ADCCROKNCFCDDSS 384  
 QY 605 PTTTGTSTSDKAPAPTTPKEAPTTPKEPAPTTPKAPATTTPETPTTSEVSTPTT 664  
 DB 385 ISGTRKWCYSTSOVAA--TKTTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 436  
 QY 665 KEPTTIHSPDESIPELSAEPTPKALENSKPEPGVTKPKPATPKEMTTAKDKTTERD 724  
 DB 437 TTTT-----TTTA-----TTTTTTTTTTTTT---KAT--- 463  
 QY 725 LRTPEPTTAAPKMTKETATTTETKESITATTTOVTSTTTOOTTPKITTLLTTLAP 784  
 DB 464 -TTTTTTTTTTP-----TTTTTKAT-----TTTTTTTTTTTTTTTAKATTATPTTPTT 510  
 QY 785 KYTTTKKITTITTEINNKKEEFAKPKDRATNSKATTPKOKOTPKAPKKTSTSKPKP---T 840  
 DB 511 TTTTAKATTITTTSGECKME---PSKRADCGYPGITTESQCKSGCCFDDSIQTKWCYYS 566  
 QY 841 MPRVRKPKTTPR 854  
 DB 567 LPQVADCKVAPSSR 580

RESULT 13  
 RPBI\_HUMAN STANDARD; PRT: 1970 AA.  
 ID RPBI\_HUMAN  
 AC P24928;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPBI).  
 GN POLR2A.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92178992; PubMed=1542581.  
 RA Witzelz M., Acker J., Vicaire S., Vigneron M., Keding C.,  
 RT "Complete sequence of the human RNA polymerase II largest subunit."  
 RL Nucleic Acids Res. 20:910-910(1992).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95347616; PubMed=7622068;  
 RA Mita K., Tsuji H., Moriyama M., Takahashi E., Neno M.,  
 RA Ichimura S., Yamuchi M., Hongo E., Hayashi A.,  
 RT "The human gene encoding the largest subunit of RNA polymerase II."  
 RL Gene 159:285-286(1995).  
 CC - FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC - CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
 CC RNA(N).  
 CC - SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC - SUBCELLULAR LOCATION: NUCLEAR.  
 CC - PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND 5.8S GENES.  
 CC - SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: X63564; CAA5125.1;  
 CC EMBL: X74874; CAA52862.1;  
 CC EMBL: X74873; CAA52862.1; JOINED.  
 CC EMBL: X74872; CAA52862.1; JOINED.  
 CC EMBL: X74871; CAA52862.1; JOINED.  
 CC EMBL: X74870; CAA52862.1; JOINED.  
 CC PIR: S21054; S21054.  
 CC MIM: 180660;  
 CC InterPro: IPR000684; RNA\_polII\_repeat.  
 CC InterPro: IPR000722; RNA\_pol\_A.  
 CC InterPro: IPR002879; RNA\_pol\_A2.  
 CC Pfam: PF00623; RNA\_pol\_A2.  
 CC Pfam: PF01854; RNA\_pol\_A2; 1.  
 CC DR PROSITE: PS00115; RNA\_pol\_II\_REPEAT: 43.  
 CC DR TRANSFERASE: DNA-directed RNA polymerase: Transcription; Zinc; Repeat;  
 CC KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 CC ZN\_FING 71 87  
 CC FT DOMAIN 1590 1958 C2H2-type (POTENTIAL).  
 CC FT CONFLICT 1067 1067 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC FT CONFLICT 1449 1449 W -> L (IN REF. 2).  
 CC FT CONFLICT 1449 1449 D -> Y (IN REF. 2).  
 SQ SEQUENCE 1970 AA: 217205 MW: 6876C25692A657E CRC64;

Query Match

8.8%; Score 503.5; DB 1; Length 1970;

Best Local Similarity 33.3%; Pred. No. 5.0e-14;  
 Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

QY 151 PPTTPKEPASTTPE-PTPTTKSAPTTKEPATTTKSAPTTKE-----PAPTTTKE 204  
 DB 1507 PSPMGISPAATPNQCATPAVGMASVSGMTPGAAGSPSASASDAGSPSPSPAWS 1566  
 QY 205 PAPTPKEPATTKEPATTTKSAPTTKEPATTTPKPAATTKEP--APTTPKEPTP 262  
 DB 1567 PTPSPSPGSP 1624  
 QY 263 TTPKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEP 322  
 DB 1625 TSPSP-SPTSPSPSPSPSPSP-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP 1671  
 QY 323 PAPTTKSAPTTKEPATTTKSAPTTKEPSPPTTKEPATTKEPATTKEPATTKEP 362  
 DB 1672 -SPTSPSPSPSPSP-SPTSPSPSPSPSP-SPTSPSP-SPTSPSP-SPTSPSP 1725  
 QY 383 KEPAATTPKEPATTTKKPAATPAKEPATTKEPATTKEPATTKEPATTKEPATTKEP 442  
 DB 1726 SY-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP 1777  
 QY 443 PTPPEELAPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTTPEEP 500  
 DB 1778 PTSP-NTSPSPSP-SPTSPSP-SPTSP-SPTSPSPSPSPSPSPSPSPSPSPSPSPSP 1828  
 QY 501 TPKETAPTPKGTAPTTLKEPATTTPK-KPAPELAPTTKEPATTKEPATTKEPATTKEPATT 559  
 DB 1829 SP-SPTSP 1875  
 QY 560 PTPKKEPATTTPKEPATTTKGAPTTLKEPATTTPKPAPELAPTTKEPATTKEPATTKEPATT 618  
 DB 1876 PTSPPT-PTTPPKY-SPTSP-SPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1922  
 QY 619 PAPTPKEPATTTPKEP--APTTPKPAATTPPTTPEPPTTSEVSTPTTKEPTTIHKSPE 676  
 DB 1923 YSPSP-SPTSP 1965  
 QY 677 STPE 680  
 DB 1966 SDEE 1969

RESULT 14  
 RPBL\_MOUSE  
 ID RPBL\_MOUSE STANDARD: PRT: 1970 AA.  
 AC P08775;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).  
 GN POLR2A OR RPO2-1 OR RPII215.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87280135; PubMed=3038894;  
 RA Ahern J.M. Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.;  
 RT "Cloning and sequence analysis of the mouse genomic locus encoding  
 RT the largest subunit of RNA polymerase II.";  
 RL J. Biol. Chem. 262:10695-10705(1987).  
 RN [2]  
 RP SEQUENCE OF 1587-1970 FROM N.A.  
 RX MEDLINE=86068017; PubMed=2999785;  
 RA Corden J.L., Cadena D.L., Ahern J.M. Jr., Dahmus M.F.;  
 RT "A unique structure at the carboxyl terminus of the largest subunit  
 RT of eukaryotic RNA polymerase II.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).  
 RN [3]  
 RP REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.

RX MEDLINE=92178992; PubMed=1542581;  
 RA Winterlith M., Acker J., Vicaire S., Vigneron M., Kedinger C.;  
 RT "Complete sequence of the human RNA polymerase II largest subunit.";  
 RL Nucleic Acids Res. 20:910-910(1992).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND 5.8S GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: M12130; AAA40071.1;  
 DR EMBL: M14101; AAA40071.1; JOINED.  
 DR PIR: A28490; A28490.  
 DR MGD: M61:98086; RPO2-1.  
 DR InterPro: IPR000684; RNA\_POL\_II\_repeat.  
 DR InterPro: IPR000722; RNA\_POL\_A.  
 DR InterPro: IPR002879; RNA\_POL\_A2.  
 DR Pfam: PF00623; RNA\_POL\_A; 1.  
 DR Pfam: PF01854; RNA\_POL\_A2; 1.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 42.  
 DR Transferrase: DNA-directed RNA polymerase: Transcription; Zinc; Repeat;  
 DR DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 FT ZN\_FING 71  
 FT DOMAIN 1590 1958  
 FT CONFLICT 1498 1498 C2H2-TYPE (POTENTIAL).  
 FT CONFLICT 1499 1498 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 FT CONFLICT 1499 1536 P->R (IN REF. 1 AND 2).  
 FT CONFLICT 1499 1536 MISSING (IN REF. 1 AND 2).  
 SQ SEQUENCE 1970 AA; 217175 MW; 7D76F38FD92A657E CRC64;

Query Match 8.7%; Score 498.5; DB 1; Length 1970;  
 Best Local Similarity 33.1%; Pred. No. 9.2e-14;  
 Matches 180; Conservative 101; Mismatches 168; Indels 95; Gaps 47;

QY 151 PPTTPKEPASTTPE-PTPTTKSAPTTKEPATTTKSAPTTKE-----PAPTTTKE 204  
 DB 1507 PSPMGISPAATPNQCATPAVGMASVSGMTPGAAGSPSASASDAGSPSPSPAWS 1566  
 QY 205 PAPTPKEPATTKEPATTTKSAPTTKEPATTTPKPAATTKEP--APTTPKEPTP 262  
 DB 1567 PTPSPSPGSP 1624  
 QY 263 TTPKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEP 322  
 DB 1625 TSPSP-SPTSPSPSPSPSP-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP 1671  
 QY 323 PAPTTKSAPTTKEPATTTKSAPTTKEPSPPTTKEPATTKEPATTKEPATTKEP 362  
 DB 1672 -SPTSPSPSPSPSP-SPTSPSPSPSPSP-SPTSPSP-SPTSPSP-SPTSPSP 1725  
 QY 383 KEPAATTPKEPATTTKKPAATPAKEPATTKEPATTKEPATTKEPATTKEPATTKEP 442  
 DB 1726 SY-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP-SPTSPSP 1777  
 QY 443 PTPPEELAPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTTPEEPPTTPEEP 500  
 DB 1778 PTSP-NTSPSPSP-SPTSPSP-SPTSP-SPTSPSPSPSPSPSPSPSPSPSPSPSPSP 1828

QY 501 TPKETAPTKGTAATLKEAPATPK-KPAKELAPTTKEPTSTSDKAPATPKGTA 559  
 Db 1829 SP-SYSTSPKKT-----PTSPSYSPSPPEVTPAPSK--YSPTSPKYSPTSPK-YS 1875  
 QY 560 PTPPKKAPATPKKAPATPKGTAATLKEAPATPKKPAKELAPTTKGT-STSIX 618  
 Db 1876 PTPSPK-SPTPPKY-SPTPSP--TSPSPVYTPTPSK-----YSPTS--PTSPSPK 1922  
 QY 619 PAPTTPKETAPTKKPP--APTTPKAPATPTTETPTTSEVSTPTTKEPTTIHKSDE 676  
 Db 1923 YSTSP-TSPSPKSTSPSTSPSPG-SPSP-----TSPSPAL-----SPD 1965  
 QY 677 STPE 680  
 Db 1966 SDEE 1969  
 RESULT 15  
 NFH\_RAT STANDARD; PRT: 831 AA.  
 AC P16884: 063368;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)  
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).  
 GN NFH OR NFH.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCBI\_TaxId:10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Brain;  
 RX MEDLINE-89065087; PubMed-3143606;  
 RA Breen K.C., Robinson P.A., Wion D., Anderton B.H.;  
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).  
 RT Identification of putative phosphorylation sites.";  
 RL FBS Lett. 241:213-218(1988).  
 RN [2]  
 RP SEQUENCE OF 37-831 FROM N.A.  
 RX MEDLINE-88309090; PubMed-2457365;  
 RA Daultigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,  
 RA Uolles P.;  
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and  
 RT in situ detection.";  
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).  
 RN [3]  
 RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.  
 RX MEDLINE-87080760; PubMed-2878828;  
 RA Robinson P.A., Wion D., Anderton B.H.;  
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide  
 RT (NF-H).";  
 RL FBS Lett. 209:203-205(1986).  
 RN [4]  
 RP SEQUENCE OF 318-831 FROM N.A.  
 RX MEDLINE-89184647; PubMed-2928342;  
 RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,  
 RA Smolowitz M., Caroll Z., Emanuel B.S., Breitner J., Rubin L.;  
 RT "Cloning of a cDNA encoding the rat high molecular weight  
 RT neurofilament peptide (NF-H): developmental and tissue expression in  
 RT the rat, and mapping of its human homologue to chromosomes 1 and  
 RT 22.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
 CC OBSERVED BY THE TWO SMALLER NF PROTEINS.  
 CC -1- THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFH IS IMPORTANT IN THE MAINTENANCE  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE

CC OF AXONAL CALIBER.  
 CC -1- PTH: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783  
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).  
 CC -----  
 CC EMBL: M37227; AAA41693.1; ALT\_FRAME.  
 CC EMBL: X13804; CAA32038.1; ALT\_FRAME.  
 CC EMBL: M21964; AAA41695.1;  
 CC EMBL: J04517; AAA41692.1;  
 CC PIR: A30796; A30796.  
 CC PIR: A25649; A25649.  
 CC PIR: B25649; B25649.  
 CC PIR: S02003; S02003.  
 CC InterPro: IPR001664; IF.  
 CC Pfam: PF00038; filament; 1.  
 CC PROSITE: PS00226; IF: 1.  
 CC Intermediate filament; Coiled coil; Neuron; Phosphorylation; Repeat.  
 CC KW NON\_TER 1  
 CC FT DOMAIN 1  
 CC FT 276 641  
 CC FT 164 164  
 CC FT 185 185  
 CC FT 193 193  
 CC FT 199 199  
 CC FT 346 346  
 CC FT 373 373  
 CC FT 482 482  
 CC FT 485 485  
 CC FT 570 571  
 CC FT 591 591  
 CC FT 727 727  
 CC FT 757 759  
 CC FT 769 769  
 CC FT 775 775  
 CC FT CONFLICT 769 769  
 CC FT CONFLICT 775 775  
 CC SO SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;  
 Query Match 8.7%; Score 497; DB 1; Length 831;  
 Best Local Similarity 27.6%; Pred. No. 5.2e-14;  
 Matches 195; Conservative 59; Mismatches 307; Indels 146; Gaps 35;  
 QY 37 PDSTTOHKNVSPKITTAKPINRPSLPNSDSKETSILVNNETVETKETTNNKQ 96  
 Db 197 PSMST--HIKVKSEKIKYV-----KSEKTEVIEEQEEIYVEEVEEED 242  
 QY 97 TSTDGKEKTSIAKETOSIEKTSKADLAPTSKYLAKTTPAETTTKCPALTTKPEP---T 152  
 Db 243 KEAQGESEEEAEAGEGAATTS-----PRAEASP---EKETSPVKEAKSPAEAKS 293  
 QY 153 PTPPKKPA-STTPKEPTPTTKSAPTTPKPA-----PTTKSAPTTPKPA-----PTTK 203  
 Db 294 PAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 352  
 QY 204 EP-----APTTPKPA-----PTTKKPA-----PTTKSAPTTPKPA-----PTTKKPA-- 245  
 Db 353 SPGAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 411  
 QY 246 --PTTKKPA-----PTTKKPE-----PTTKKPAAPTTPKPE-----PTTKKPA-- 287  
 Db 412 KSPYEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAE 470  
 QY 288 PKKAPPTTPKPAAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 338

```
Db      471 AKSPAEEA--KSPAGAKKKPAEAEKSAPAEAKSPPAEAKSPAEAKSPAEAKSPVEYKSEPERAKSP 528
Oy      339 APTTKS-----APTTPKEBPSPTTKEBA-----DTTPKEBAPTTPKKPADTTPEKBAPT 389
Db      539 VKEGAKSLAEAKSBEKASPKASPBEEKIBPAAEVKSEKAKSPMRKEAKSPERAKTIDVKSPE 588
Oy      390 PKEPAPTTPKKPAPTAPAREBAPTTPKETAPTTPKLLPTTPPEKLAAPTTPPKAPAPTPEEL 449
Db      589 AKRPAKEAKKRPAIDRSQEQVSKAPEAKSPEKEERT--EKVAPKEEVKSPV--EEV 644
Oy      450 APTPEEBPTPTTPPEBPATTPKAAPNTPEKBAPTTPKEBAPTTPKEBAPTTPKETAPTTP 509
Db      645 ---KAKEBPKEVEEBKTPTATKTEVEKSSKKDEAPKEAQK-KAEKEKPLETRKDSGEA 700
Oy      510 PKGTAPTLPKEBAPTTPPKKPAPKELAPTTTKEPTTSODKPAATTTPKGTAAPTTPKEBAPT 569
Db      701 KKEEA-----KKKKAABPEETPAKLGVKYEAKPEKAEDKA-----KKBSAP 744
Oy      570 TPKEBAPTTPKGTAPTTLKEBAPTTPPKKPAPKELAPTTTGPTTSODKPAAPTTPKETAP 629
Db      745 SEKE-----KPK-----KEEVPAPEKKDKTE-----EXTTESKKREKKPMEMAK 784
Oy      630 TTPKEBAPTTPKKRAPTTTPETPTTPNSEVTPTTTEKEPTTIHNSPOE 676
Db      785 AKKED-----KGLPOEPSKPTTERKAKSSSTDOAKSDOSSEKAPBD 824
```

Search completed: April 26, 2002, 16:19:50  
Job time: 375 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:17:05 ; Search time 49.78 seconds  
(without alignments)  
480,986 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1263

Perfect score: 5698

Sequence: 1 VKDKKKRKKKKPPKPPVY.....VAALSTAKYKNWPEYFFK 1064

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/CTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfilest.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	522	9.2	805 4 US-09-103-429A-4	Sequence 4, Appli
2	508	8.9	1837 3 US-08-928-361B-5	Sequence 5, Appli
3	506.5	8.9	744 6 520236-25	Patent No. 520236
4	498.5	8.7	786 4 US-09-103-429A-3	Sequence 3, Appli
5	489	8.6	1721 3 US-08-700-651-5	Sequence 5, Appli
6	488.5	8.6	1721 3 US-08-928-361B-6	Sequence 6, Appli
7	488.5	8.6	826 1 US-07-638-431-2	Sequence 2, Appli
8	476.5	8.6	826 5 PCT-US92-00018-2	Patent No. 520236
9	476.5	8.4	652 6 520236-13	Sequence 5, Appli
10	452	7.9	960 4 US-09-219-849-5	Sequence 5, Appli
11	424.5	7.4	1867 2 US-08-479-537A-5	Sequence 5, Appli
12	424.5	7.4	1867 4 US-08-083-116-5	Sequence 5, Appli
13	424.5	7.4	2035 2 US-08-479-537A-2	Sequence 2, Appli
14	424.5	7.4	2035 4 US-09-083-116-2	Sequence 2, Appli
15	419.5	7.4	2476 2 US-08-276-967-2	Sequence 2, Appli
16	419	7.4	1185 4 US-09-041-886-23	Sequence 2, Appli
17	417	7.3	829 1 US-08-642-255-132	Sequence 132, App
18	417	7.3	829 1 US-08-397-633A-53	Sequence 53, Appli
19	417	7.3	837 1 US-08-175-155-68	Sequence 68, Appli
20	417	7.3	837 1 US-08-477-509B-103	Sequence 103, App
21	417	7.3	837 1 US-08-642-255-101	Sequence 101, App
22	417	7.3	837 3 US-08-707-237A-75	Sequence 75, Appli
23	417	7.3	837 3 US-08-482-085B-103	Sequence 103, Appli
24	417	7.3	897 1 US-08-397-633A-50	Sequence 50, Appli
25	413.5	7.3	907 3 US-08-783-774-2	Sequence 19, Appli
26	413.5	7.3	907 5 PCT-US95-04611A-19	Sequence 65, Appli
27	404.5	7.1	408 1 US-07-609-716-65	Sequence 65, Appli

## ALIGNMENTS

28	404.5	7.1	408 4 US-08-475-411A-65	Sequence 65, Appli
29	404.5	7.1	408 4 US-08-478-029A-65	Sequence 65, Appli
30	398.5	7.0	682 1 US-08-642-255-126	Sequence 126, App
31	398.5	7.0	682 1 US-08-397-633A-36	Sequence 36, Appli
32	381	6.7	1848 4 US-08-296-791-6	Sequence 6, Appli
33	381	6.7	1848 5 PCT-US95-10661A-6	Sequence 6, Appli
34	376	6.6	960 4 US-09-219-849-6	Sequence 6, Appli
35	374.5	6.6	1537 1 US-08-325-267A-2	Sequence 2, Appli
36	373	6.5	1231 3 US-08-904-263A-4	Sequence 4, Appli
37	371	6.5	761 2 US-08-707-237A-84	Sequence 84, Appli
38	371	6.5	762 1 US-08-642-255-114	Sequence 114, App
39	371	6.5	762 1 US-08-397-633A-26	Sequence 26, Appli
40	370.5	6.5	1064 1 US-08-642-255-62	Sequence 62, Appli
41	368	6.5	762 1 US-08-642-255-120	Sequence 120, App
42	368	6.5	762 1 US-08-397-633A-31	Sequence 31, Appli
43	367.5	6.4	1187 1 US-08-320-559-28	Sequence 28, Appli
44	367.5	6.4	1187 3 US-08-545-860D-28	Sequence 28, Appli
45	367.5	6.4	1187 5 PCT-US94-04496-28	Sequence 28, Appli

RESULT 1  
US-09-103-429A-4  
Sequence 4, Application US/09103429A  
Patent No. 6187558  
GENERAL INFORMATION:  
APPLICANT: Granados, Robert R  
APPLICANT: Wang, Ping  
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
TITLE OF INVENTION: C DNA and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
STREET: 118 No. 6187558th Tlaga  
CITY: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
\* CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/103,429A  
FILING DATE: 24-JUN-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 805 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-103-429A-4  
Query Match 9.2%; Score 522; DB 4; Length 805;  
Best Local Similarity 27.8%; Pred. No. 76-27;

[illegible]

```

?      CITY: Palo Alto
?
?      STATE: CA
?
?      COUNTRY: USA
?
?      ZIP: 94306-1840
?
?      COMPUTER READABLE FORM:
?
?      MEDIUM TYPE: Floppy disk
?
?      COMPUTER: IBM PC compatible
?
?      OPERATING SYSTEM: PC-DOS/MS-DOS
?
?      SOFTWARE: PatentIn Release #1.0, Version #1.30
?
?      CURRENT APPLICATION DATA:
?
?      APPLICATION NUMBER: US/08/928,361B
?
?      FILING DATE: 12-SEP-1997
?
?      CLASSIFICATION:
?
?      PRIOR APPLICATION DATA:
?
?      APPLICATION NUMBER: US 60/026,062
?
?      FILING DATE: 13-SEP-1996
?
?      ATTORNEY/AGENT INFORMATION:
?
?      NAME: Verny, Hana
?
?      REGISTRATION NUMBER: 30,518
?
?      REFERENCE/DOCKET NUMBER: 480,76-1(HV)
?
?      TELECOMMUNICATION INFORMATION:
?
?      TELEPHONE: 650-324-1677
?
?      TELEFAX: 650-324-1678
?
?      INFORMATION FOR SEQ ID NO: 5:
?
?      SEQUENCE CHARACTERISTICS:
?
?      LENGTH: 1837 amino acids
?
?      TYPE: amino acid
?
?      STRANDEDNESS:
?
?      TOPOLOGY: linear
?
?      MOLECULE TYPE: protein
?
US-08-928-361B-5

```

```

Query Match          8.9% ; Score 508; DB 3; Length 1837;
Best Local Similarity 21.3%; Pred No. 1,5e-25;
Matches 295; Conservative 100; Mismatches 496; Indels 496; Gaps
48

QY      33 KVTTPDITSTQHNVKSPKITTAKEDINRPSLPNSDSIKETSITYNKEETVETKETTT 92
DB      307 KHITTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 366
QY      93 TNKOTSIDGKEKKTSAKETOSIEKTSNKDLAPISKYLAKPPTKATITTKGALTPKEPT 152
DB      367 TTTTTTTTTKKPPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 426
QY      153 PTTPEKPASTPKRPPTTIKASAPTTPKEP--APTTPSAPTPEKAPTTPKEAPTTP 210
DB      427 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 486
QY      211 KEAPTPTTKEBAPTTPKSAPTTPKEBAPTTPKKKAPPTTKEBAPTTPKE----- 259
DB      487 TTTATTITTKKTTTTTTTTTTTTTTTTTTTTTATTTTTTSETESVIKPDGMGLEKNGECE 546
QY      260 -----PPPTT-----PK 266
DB      547 AKGATTVGVIGDKRIENGMAFTWIPNDTHVRFRKVADVGNLTISVRCGAGAKLEFPD 606
QY      267 EKAPTTPK-----APTPEKA-PTAPKKP-----291
DB      607 RSLDFITPVAGHNSCSITIVSGGGKIHVSPGSDVLSIAPIQPCELFWNYCDICT 666
QY      292 -----APTTPKEBAPTTP--PKEBAPTTPKEBS-PTPPKEBAPTTPKS- 330
DB      667 AKYGAIHSGYQTADSEVVITTTAKPTTTTGAPQPPTTTTGSFSKPTTTTTTKATTTTTI 726
QY      331 ---APTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPPKKAPPTTPKPEA 386
DB      727 LNPITTTITTOKRPTTTTTTKVGPGRPIATTTTTPKPIVTTTTTKATTTTTVPTI--- 781
QY      387 PTTPEKPAPTTPKKAP-----TAPEKAP-----411
DB      782 TTTTRKDEMTTTTTPLPDIIGDIEITLPIEKMLDKTIRMIIDYNSGLILDSNDEFIPSSO 841

```

QY 412 -----TTPKETA-PTTP-----KKL 425  
Db 842 AGOIAIDTSLNLEPVQTHKSTGLIPDMVGLPEDPKSGNLVHYTNQTMGSLSVSLAAKNL 901  
QY 426 T-----PTTPKELAPTP-----439  
Db 902 TVDIDETVYGLPIDLTGLYPLDPVSLIPNPETGELFDPISDEIMNGTIAGIVSGISASES 961  
QY 440 ----KPAPTTPEE-----LAPTT-----PEEPTPTP--EAPPTPKAA 474  
Db 962 LLSQKSAIPDPAITMVMYGEFGLNIPATGMIGSLGSPSQTPSPSEIDEGCIIPEPVAA 1021  
QY 475 PNPKEAPPTPKKEPATTTKE-----PA 498  
Db 1022 ANMDKEFLSIPSVESIPKEDQKIDSISELMYDIESGRLIGOVSKRPIGSIAGDLNPI 1081  
QY 499 PTPPKETAPPTPKGTATTTTKEEAPATTPPKKPAKELAPTTKEPTSTSGKRAPPTPKGT 558  
Db 1082 MKTPTOTDSVTKRPIDTT--GLPFP--PTGHLINPTNNMTMDSFAGAYVAVSNGI 1136  
QY 559 APPTT-----PKRAPPTPKKEPATTPKGTATTTKEAPATTP-----KKPAK-- 601  
Db 1137 KTDNVYGLPVDETITGLPKDVSDFPNSITGELVDSPTGKPINNYAGIYSGKRGLEPIE 1196  
QY 602 ----ELAPTTKGP-----TST-----SDKRAPTTKEETAPTTKEPA-- 636  
Db 1197 DENGNLFDSTKLPIDGNQNLVNPETNSVSGSTGKRPGLIPVNGGQVVDDEAKDQ 1256  
QY 637 -----PTTPKRAPPTPEPTTSEVSTPTTKE-PTTIHKS--PDESTPEL 681  
Db 1257 ADKGDGLIAPPTNSINKDQVNTQYSNTGGINETGKVIPLGSLNIPSEFTPOQ 1316  
QY 682 SAAPTAKALE--NSPKREGVPTTKTPAATKPEMTTAKDKTTERDLKTTPETTTAARK 738  
Db 1317 TDEITKAPDVTYGLPYDPSGTGLIDPATKLPICGSVAGDEILTEVLNITTDVETGLP-I 1375  
QY 739 TKEATATTEKTESKITATTTQVSTTQDTPTPEKITITLIKTTLAKVTTTKITITTE- 797  
Db 1376 DLETGILPRDP-----VSLGFLQPLNCTLYD-----PSNKKRPIGSHS 1411  
QY 798 --IMNKPEETAKPKDRATNSKATTPKQKP--TKAP--KKPTSKKPTWPRVAKPKT-- 849  
Db 1412 GFNGTSGQSHKDPSTG-----KPLDPNTGLPDEDESGSLINETDKIQGSISGTF 1465  
QY 850 TTPPKMTSTMBELNPTSLIAEAMLOTTTRPQNTPSKILVEVNPK-SEDA----- 898  
Db 1466 MEVPKPGCEGNGIMTPELIAL-----NKLPTSNENVNISPRSSDAVDRPTNTW 1518  
QY 899 -----GGAGETPHMLT-----RPHVMPVPTPDMOYLPRVNOGIII 936  
Db 1519 NKISGTFQVQDKTIPGSAASVITHALCTPTQIDPTTGLP--SDPSTGLPPIGPNVLV 1576  
QY 937 NMLSD-----ETNIC-----NCKPVDGLITL--RNGTLV-AF 966  
Db 1577 DQDTEQIKSVYVSLYVKEKNIVTEAAYGLVPDKTGFPIPISTYLPFAKNGELIDPI 1636  
QY 967 RGHYF 971  
Db 1637 SGKYF 1641

RESULT 3  
5202236-25  
; Patent No. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSERG,  
; SUSAN L.; MCCARDISS, RUSS; WEI, TERA; PILULA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
; PROTEIN  
; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,762  
; FILING DATE: 25-MAY-1990  
; APPLICATION NUMBER: 82,456

; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945  
; FILING DATE: 24-NOV-1986  
; APPLICATION NUMBER: 650,128  
; FILING DATE: 13-SEP-1984  
; SEQ ID NO: 25  
; LENGTH: 744  
5202236-25

Query Match 8.9%; Score 506.5; DB 6; Length 744;  
Best Local Similarity 29.9%; Pred. No. 6,7e+26;  
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 51 PKTTTAKPINRPSLPNSDTSKESLTVNKEETVETKETTNTNQTSDGKEKTSKAKE 110  
Db 24 PKMTYPTTKPKRPSYP-----YKSKPTLY-----KPKIT----- 53  
QY 111 TOSIEKSAKDAPTSKVLAKPTPKAETTTKGPALTTKEEPTTPKE-----PASTTKE 166  
Db 54 -----YPTTK--AKPS-----YPTTKPKKTYPTTKKLYRPTTKPK 92  
QY 167 PPTTKISAPT--TPKEPATTTKSAPTTPKKEPATTTTKEPATTPKKEPATTTKKEPA--P 223  
Db 93 SYPTTKSKPTKPKKITYPPTYKAKPSYPTTKPKKKTTPPT--YKRLITYPTTKKASYP 151  
QY 224 TTTKSAPTTPKKEPATTTKPKAPT--TPKEPATTT--PKEPTTPKKEPATTTKEP-- 275  
Db 152 PTYKPKPSYP--PSYKTKTTPPTPKKLTYPPTYKPKPSYPSYKPKTYPPTYKPKLT 209  
QY 276 -APTTKKEPA--PTAPKKA--PTTPKKEPA--PTTPKKEPA--PTTKKES--PTTPKKEPA 324  
Db 210 YPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 269  
QY 325 -PTTKSAPTTTKKEPATTTKSAPTTPKESPTTTKKEPA--PTTPKKEPA--PTTPKPK 377  
Db 270 YPTTKAKPSYPT--YPTTKAKPSYPT--PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 322  
QY 378 A--PTTPKKEPA--PTTPKKEPA--PTTPKKEPA--PTAPKKEPATTTKEPATTTKPKKLTPTT 429  
Db 323 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 382  
QY 430 PEKLAPTTPE--KPAATPEELAPTTPEEPT--PTTPEEPA--PTTPKA--AAPNT--P 478  
Db 383 TYKAKPSYPTTKAKPSYPT--PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 438  
QY 479 KEPATTTKEEAPPTT--PKRAPPTTPKEPATTT--PKGTAPTTLKEPATTPKPKAPK 532  
Db 439 SYPTTKPKISYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 498  
QY 533 ELAPTTKEPTSTSDKPAPTTPKGTATTPKKEPATTTKKEPATTTKGTATTLKKEPA- 591  
Db 499 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 546  
QY 552 -PTTPKKAPELAPTTTKGPT--STSDKPAPTTPKKEPATTTKKEPATTTKPKKA--PT 646  
Db 547 PPTTKAKPTYK--AKPTTKAKPTTSTYTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 601  
QY 647 --TPPTPTTSEVSTPTTKE-----PTTIHKSPPD-----ESTPELSAEP--PKALENS 693  
Db 602 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 661  
QY 694 PKREGVPTTKPAA-----TKPEMTTAKDKTTERDLRTTPEETTAAP-----KMT 739  
Db 662 PSYP--PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 719  
QY 740 KETATTEKTESKITATTT 759  
Db 720 YKAKPTTSTYKAKPTYPST 739

RESULT 4  
US-09-103-429A-3

Sequence 3, Application US/09103429A  
Patent No. 6187558.  
GENERAL INFORMATION:  
APPLICANT: Granados, Robert R  
APPLICANT: Wang, Ping  
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
TITLE OF INVENTION: cDNA and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
STREET: 118 No. 6187558th Tloga  
CITY: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/103,429A  
FILING DATE: 24-JUN-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 786 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEtical: NO  
AMTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-103-429A-3

Query Match 8.7%; Score 498.5; DB 4; Length 786;  
Best Local Similarity 29.1%; Pred. No. 2.4e-25;  
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;  
QY 132 PTPRAETTTGPAITTKREPTTPPKREPASTTPKEPTTTIKSAPTTPKREPAITTKSAP 191  
Db 93 PGPRAE-ITQAPATTT---QAPITTOAPITTTT-----QAPITTOQ-ATTTTQAP 135  
QY 192 TTPREPAATTTKEPAATTTKEPAATTTKSAPTTPKAPATTPKAPATTPKE 251  
Db 136 ITTQ-ATTTQ-ATTT---QAPITTTQ-APITTTQAPITTTQ-APITTTQ-APITTTQ- 183  
QY 252 PAPTTPKEPTTPKREPAATTPKREPAATTPKAPATPKAPATTPKREPAATTP-KEPA- 308  
Db 184 -APTITQAPITTTQATTPATTPATTPATTPATTPAA--TTPATTPGVAPEAPASAVWPIC 241  
QY 309 -----PTTTEPSPTTPK- 321  
Db 242 EILPNCBPADFDHLLIPHDKCYCNLFYQCSNGYTFEORCPBGLYFNPIYVOCDSBANVEC 301  
QY 322 ---EPAPITTT----- 328  
Db 302 DGEISAPAPVYEGNEDDEDIDGDLNCGPANFEIDWLLPHGNCDKXYQCCHGLVERR 361  
QY 329 -----KSAPTTPKEPATTTKSAPTTPKREPSPTTTPKEP 361  
Db 362 CGAGTHFSFELQCDHIEIVGCTLPGESEEVVDVDEDA-CTGWCYCPTEPIEMERLPGNCP 420

QY 362 APTT-----PKP-----APTTPKKAPPT----- 381  
Db 421 ADFSIDHLLPHESDCOGYLOQCVHGQTARPCGNLHSPATQSCSESVTAGCOVFECDSD 480  
QY 382 ---KREPAATTPKREPAATTTKPPATPKREPAATTPKREPAATTPKREPAATTP 438  
Db 481 NOSTSTAPATPAAPTAAPTAAPTAAPTA-APSTVVPATTPATAPAPPTT---AITT- 533  
QY 439 EKPAATPEEIAAPTTPEEPTTPPEEPAATTPKKAAPT---PKREPAATTPKREPAATTPKE 496  
Db 534 -PATTAAPTAAPTAAPESPTVTVP-PTAAPTAAPTTAPEIPIVTV---SAPTAPPT 587  
QY 497 PAPTTPKETAPITTPKGTAPITLKEPAATTPKPPAPKELAPITTEPTSTSDKAPATTPK 556  
Db 588 AAPTAAPTAAPTTAPEIPIVTVSPTTAAPTAAAP---APNTT-----VIVPP 632  
QY 557 GTAPTTPKREPAATTPKREPAATTPKGTAPITLKEPAATTPKPPAPKELAPITTKGPTSTS 616  
Db 633 TAAPTT-AAAPANTTVVYPTTAAPTAAAPVVAH-----APNTTAAVPTTS 677  
QY 617 DKPAATTPKETAPITTPKREPAATTPKPP 643  
Db 678 ---APATTPEDDIDP--PLPNDPINP 699

RESULT 5  
US-08-700-651-5  
Sequence 5, Application US/08700651B  
Patent No. 6015882  
GENERAL INFORMATION:  
APPLICANT: PETERSEN, CAROLYN  
APPLICANT: LEECH, JAMES  
APPLICANT: NELSON, RICHARD, C.  
APPLICANT: GUT, JIRI  
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
FILE REFERENCE: 480.19-4(HV)  
CURRENT APPLICATION NUMBER: US/08/700,651B  
CURRENT FILING DATE: 1997-08-14  
EARLIER APPLICATION NUMBER: 08/415,751  
EARLIER FILING DATE: 1995-04-03  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 1721  
TYPE: PRT  
ORGANISM: Cryptosporidium parvum  
US-08-700-651-5

Query Match 8.6%; Score 489; DB 3; Length 1721;  
Best Local Similarity 22.1%; Pred. No. 2.4e-24;  
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;  
QY 5 KKNRTKKKPPKPPVYDEAGSGLDNGDFKYTPDSTTOHNKYSIPKITTAKPINPRS 64  
Db 116 RSNMTK---ETEPSANYAGVRSN-ETKTTPEPSANT---NFLVDPKI-----N 158  
QY 65 LPNDSJTSKE-----TSLVNKEETVEKTKETTTNQTSDGKEKTSK 109  
Db 159 APCSENSFEOGOIFDMGSKVYIPYTKCVGKHHTTTTTTTTTTTTTTTTTT 208  
QY 110 ETOSIEKTSAKDLAPTSKVLAKPTPKAETTTTKGPAITTPKEPTTPKREPAATTPKEPPT 169  
Db 209 -----TT 240  
QY 170 TTIKSAPTTPKREPAATTTKSAPTTPKREPAATTTKREPAATTPKREPAATTTKSA 229  
Db 241 TTTTNTT 300  
QY 230 PTTKREPAATTPKPPATTTTPKREPAATTPKREPAATTPKREPAATTPKREPAATPAK 289

[illegible]

```

US-08-928-361B-6      ; Sequence 6, Application US/08928361B
US-08-928-361B-6      ; Patent No. 6071518
US-08-928-361B-6      ; GENERAL INFORMATION:
US-08-928-361B-6      ; APPLICANT: Petersen, Carolyn
US-08-928-361B-6      ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
US-08-928-361B-6      ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS AND FRAGMENTS
US-08-928-361B-6      ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
US-08-928-361B-6      ; TITLE OF INVENTION: SPECIES INFECTIONS
US-08-928-361B-6      ; NUMBER OF SEQUENCES: 30
US-08-928-361B-6      ; CORRESPONDENCE ADDRESS:
US-08-928-361B-6      ; ADDRESSEE: PETERS, VERNY, JONES & BIRSA
US-08-928-361B-6      ; STREET: 385 Sherman Avenue, Suite 6
US-08-928-361B-6      ; CITY: Palo Alto
US-08-928-361B-6      ; STATE: CA
US-08-928-361B-6      ; COUNTRY: USA
US-08-928-361B-6      ; ZIP: 94306-1840
US-08-928-361B-6      ; COMPUTER READABLE FORM:
US-08-928-361B-6      ; MEDIUM TYPE: Floppy disk
US-08-928-361B-6      ; COMPUTER: IBM PC compatible
US-08-928-361B-6      ; OPERATING SYSTEM: PC-DOS/MS-DOS
US-08-928-361B-6      ; SOFTWARE: Patentin Release #1.0, Version #1.30
US-08-928-361B-6      ; CURRENT APPLICATION DATA:
US-08-928-361B-6      ; APPLICATION NUMBER: US/08/928,361B
US-08-928-361B-6      ; FILING DATE: 12-SEP-1997
US-08-928-361B-6      ; CLASSIFICATION:
US-08-928-361B-6      ; PRIOR APPLICATION DATA:
US-08-928-361B-6      ; APPLICATION NUMBER: US 60/026,062
US-08-928-361B-6      ; FILING DATE: 13-SEP-1996
US-08-928-361B-6      ; ATTORNEY/AGENT INFORMATION:
US-08-928-361B-6      ; NAME: VERNY, Hana
US-08-928-361B-6      ; REGISTRATION NUMBER: 30,518
US-08-928-361B-6      ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
US-08-928-361B-6      ; TELECOMMUNICATION INFORMATION:
US-08-928-361B-6      ; TELEPHONE: 650-324-1677
US-08-928-361B-6      ; TELEFAX: 650-324-1678
US-08-928-361B-6      ; INFORMATION FOR SEQ ID NO: 6:
US-08-928-361B-6      ; SEQUENCE CHARACTERISTICS:
US-08-928-361B-6      ; LENGTH: 1721 amino acids
US-08-928-361B-6      ; TYPE: amino acid
US-08-928-361B-6      ; STRANDEDNESS:
US-08-928-361B-6      ; TOPOLOGY: linear
US-08-928-361B-6      ; MOLECULE TYPE: protein
US-08-928-361B-6

```

```

Query Match 8.6%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 2,4e-24;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 411

QY 5 KKNRTKKKPPKPPVVDGASGLDNGDFVTTTPDDSTTGHNVSTSPKTTAKPIIPRRS 64
Db 116 RSNETK---TTEPSANTYAGVYRSN-EKTEPPSANT---NFLVDPKI-----N 158
QY 65 LPNNDYTSKE-----TSLVNKEVTVETKEVTTTNKQSTGDEKERTTSAK 109
Db 159 APCSENSEFEGQIFDMGSKVIPIYTKCVGVKHTTTTTTTTTTTTTTTTT 208
QY 110 ETQSIETKSADLAPTSKVLKPPKPAETTTGKGLATTPKKEPTPTTPKEPASTTPKEPP 169
Db 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
QY 170 TTIKSAPTPPKEPAPTTTTSAPTPPKKEBPATTPPKKEBPATTPPKKEBPATTTKSA 229
Db 241 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
QY 230 PTPKKEPAPTPPKPAPTPPKKEBPATTPKKEPPTTPPKKEBPATTPPKKEBPATPAK 289
Db 301 TTTTKKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 350
QY 290 KPAETPKEPAPTPPKKEBPATTTKEBSPPTPKKEBPATTTKSAPTTTKKAPPTTKSAPT 349
Db 351 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 406

```





OY	377	PA--PTTPKEBA--PTTPKEBAPTITTKKAPATAPKKEAPATPREKAPTPTKTLPTTPK	432
Db	309	PSYPTTKAKSPSYPTTKAKSPSYPTTKAKSPY---PTYKAKSPSYPTTKAKSP--YK	362
OY	433	LAPTPPE--KPAPTPEELAPTPPEEPT--PTTPEEBA--PTTKAAPTPTKEBAPTTPK	487
Db	363	AKPTIYPSYTKAKSPY---PTYKAKSPSYPTTKAKSPSYPTTK--AKSPY---PTYKA	413
OY	488	EPA--PTT---PKEBAPTTPKEAPTT--PKGAPTTLKEBAPTTPKKAPKELAPTTPK	540
Db	414	KPTYSTYKAKSPSYPTTKAPISIPPTYKAKSPSYTYKAKSSYPTTKAKSPSYPTTKA	473
OY	541	EPTSTISOKRAPPTPKGAPTTPKEBAPTTPKKEBAPTTPGATLTLKEBA--PTTPKKP	598
Db	474	KPTIYPSYTKAPTYK--AKSPY---PTYKAKS-----YPTTKAKSPSYPTTKAKP	521
OY	599	APKELAPPTTKGPTSTUSDKAPAPTPEKETAAPTTPKEBA--PTTPKKPA--PTT	651
Db	522	TYK--AKPTTKAKSPSYPTTKAKSPY---PTYKAKSPSYPTTKAKPTSYTYKAKSPY	576
OY	652	PPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPPTPKALENSPKRECVPTTKTPAATKE	711
Db	577	PTYKAKSPSYPT-----YKAKSPSYPTTKAKPTYKA-----KPTYPSI---YKAKS	620
OY	712	MTTTPAKOTTERDLKTTPETTAAP	736
Db	621	YPTTHKAPPTYKAKSPYPTTKAKP	645

```

RESULT 10
US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELDE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HAIR EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; US-09-219-849-5

```

Query Match	7.9%;	Score 452;	DB 4;	Length 960;
Best Local Similarity	25.6%;	Pred. No. 3.3e-22;		
Matches 173;	Conservative 60;	Mismatches 309;	Indels 134;	Gaps 41.

QY 132 PTPKAEITTKGPAITTKPEPTTPKBPASNTPEPTTPKASPTTKBPAP--PTTKS 189  
Db 72 PGPAPCPGSRDGP--PGAPAPGAP--PGSRDPPPGAPG-PAPPGPSRDPGGPAPGPA 126  
QY 190 APPTPKBPAPTTTKEPA-PTTPKBPAPTTTKEPAPTTKASPTTKBPAPTTT---KKPA 245  
Db 127 GPPGSRDGGPGAPCPAPGPPGSRDGGPAPGPPGAPGAPGPPGSGSNDPG 186

QY	246	PTTKEAPTTT---	KEPTPTTKEPA	-	PTTKEAPTTTKEPA	-----	284					
Db	187	P-	PGAGPGAPGPGSRDPGGAGPGAPGPGSRDPPGAGHGAAGKGAHGAAGPGAGAG				244					
QY	285	-----	PTAPK-KEAPTTT---	KEAPPTTKEPA	-	PTTKEPSPTTKEAPTTTKS	330					
Db	245	PAGKGAHGAAGPGAGPGAPGPGSRDPGGAGAGGAPGPGSRDPG-	PGAGP	----	A		298					
QY	331	APTTTKEBAPTTT	SAPTTKEPSPTTKEBAPTTTKEBAPTTT---	KKPAPTPTTKEBAP			387					
Db	299	GPGGSRDGP-	-----	PGAGPGAPGPGSRDGP-	PGAPGAPGPGSRDGP-	PGAPG	348					
QY	388	TTT---	KEBAPTTT	KKPA	-----	PTAPKEBAPTTTKEBAPTTT	PKLTPTTEKLA	437				
Db	349	AGPPGSRDPGGAGPGAPGPGSRDPPGAPGAPGPGSRDGP-	PGAGGAPGPG				404					
QY	438	PEKAPPT-TPEELAPTTT	EEBPTPTTKEBAPTTT	PKAA	-	PNTKEBAPTTT---	KEBAPT	492				
Db	405	SRDPPPGAPGAPGPGSRDGP-	PGAGPGAPGPGSRDPPGAPGAPGPGSRDPP				462					
QY	493	TPKEBA	-----	PTTKEET	-	APTTPKGT	-----	APTTLKEBAPTTT	KKP	-	APKELA	535
Db	463	GAHGAAGKGAHGAAGPGAGPGKGAHGAAGPGAPGAPGPGSRDGP	PGAPGAG				522					
QY	536	PTTTEPTSTSDKAPTTT	KGTA	-	PTTKEBAPTTT	---	KEBAPTTT	KGTA	-	PTTLEK	590	
Db	523	PBGSRDPPGGA	-	PGAPGPGSRDPPGPGAGGAPGPGSRDPPGAPGAPGPGSRD			580					
QY	591	APTTTKEBAP	-----	KELA	PTTTKGPTST	-	TSDKAPTTT	PKETA	PTTKEBAPTTT	PKK	642	
Db	581	GP-	-	PGAGPGAPGPGSRDPPGAGGAPGPGSRDPPGAPGAPGPGSRDPPG-	PGA		636					
QY	643	PAPTTP	---	ETPTTSEVSTPTT	TKAPTTHKSPDESTELSAEPT	PALEN	-----	S	693			
Db	637	PGAPGPGSRDPPGAGPGAPGAPGPGSRDPPG-	PGAPGAPGPGSRDPPGAPGAPGPG				695					
QY	694	PKBPGPTT	TKPATK				709					
Db	696	SRDPPGAGHGAAGK					711					

RESULT 11  
 US-08-479-537A-5  
 Sequence 5, Application US/08479537A  
 Patent No. 5861381  
 GENERAL INFORMATION:  
 APPLICANT: CHAMSON, Pierre  
 APPLICANT: KIENY, Marie-Paule  
 APPLICANT: LATHIEU, Richard  
 APPLICANT: HAREVEUET, Mara  
 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
 TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
 STREET: P.O. Box 1404  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/479,537A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: FR 90/13101  
 FILING DATE: 23-OCT-1990  
 PRIORITY APPLICATION DATA:



```

APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Thr or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-5

```

Query Match 7.4%; Score 424.5; DB 2; Length 1867;  
Best Local Similarity 23.7%; Pred. No. 4.5e-20;  
Matches 253; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

```

75 TSLTVKETTVEKETTNTKOTSDGKEKTSAKETOSIEKTSKD-LAPTSKVLAKPT 133
16 TVLT-----VTGSHASSTPGGKETSAIORSSVPSSTKNAVSMSSVLSHS 65
QY 134 P-KAETTTGCP--ALTPREP-----TPTTPKEPASTTPEKPTTIKSNP 176
DB 66 PGSGSTTGGQVTLAPATEPASGAATWGDVTSVPTRPALGSTTPPAHVT---SAP 122
QY 177 TTPKEPAPTT-----KSAPTTPKEP-----APTTPKEPAPTTTPKEPAPT 217
DB 123 --DNKPAEGSTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVT 180
QY 218 TKE--PAPTTT-----KSAPTTPKEPAPTTTPK-----KPAPTTPKEPAPTTTPKPTPTT 264

```

```

DB 181 APDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVT 240
QY 265 PKE--PAPTTKEP-----APTTPKEPAPTAK-----KPAPTTPKEPAPTTPKEPAPT 311
DB 241 APDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVT 300
QY 312 TKEBPTTPKEPAPT--TKSAPTTPKEPAPT-----TKSAPTTPKEPAPTTPK-----E 360
DB 301 APDXRP--XPGSTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVT 359
QY 361 PAPTTTPKEPAPTTPK-----KPAPTTPKEPAPTTPK-----EPAPTTPKKAAPTAPKEPAPT 412
DB 360 SAPDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVT 419
QY 413 TPKETAP---TTPKRLTPTTPKEKLAPTTPKEKAPPTTPPELAPTTPEEPTTPTEPAPT 468
DB 420 SAPDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVT 471
QY 469 TPKA---AAPNTPKKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 523
DB 472 APXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGT 530
QY 524 TTP-----KKPAPELAPTTTKEPTSTTSOKPAP--TTPKGTAPTTTPKEPAPTTP 571
DB 531 TAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGT 585
QY 572 KEPAPTTPKG---TAPTTLKEPAPTTP-----KKPAPELAPTTTGGTSTTSD 617
DB 586 PAXGTAAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGT 644
QY 618 KPAP--TTPK-----ETAPTTTPKEPAPTTP-----KKPAPTTPPTTPPTSEVSTP 661
DB 645 RXPXGTAAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGT 702
QY 662 TTPKEPTT-----IHKSPESTPELSAEPTRKALENSPKRPGVPTTKTPA----- 707
DB 703 DXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGT 762
QY 708 -TKPEKTTTAKK---TTERDLRTTPETTTPAAMKTKETATTTETKTSKITATTTQVTS 763
DB 763 DXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGT 820
QY 764 TTTQDTPPEKITTLLTTLAPRVTTTKTITTTTELMMKPEELAKKDKATTSKATTPRQ 823
DB 821 APDXRPXG---STAPXAHGVT-----SAPDXRPXGSGTAPXAHGVTSAFDXRPXGSGT 868
QY 824 KPTKAPKRPSTTKKPKTPRVAKRPKTPTPPKMTSTMPLELNPSTRIADAMLQTTTRPQ 883
DB 869 -GSTAPXAHGVTSAFDXRP--XPGSTAPXAHGVTSA--PDXRPX-----Pest 911
QY 884 -PNSKLVEVNKSEEDGAGAEGETPMMLLRPHVMEVTPDMQDYLPRVNOGIIINPMLSD 942
DB 912 APXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVT 957
QY 943 ETNINCKRPVVDLIT--LNGTLVAFRGHYFMMLSPFSPSPARRTTEWGIPLSPID 997
DB 998 VTSAPDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFDXRPXGSGTAPXAHGVTSAFD 1003

```

RESULT 12  
US-09-083-116-5  
Sequence 5, Application US/09083116  
Patent No. 6203795  
GENERAL INFORMATION:  
APPLICANT: CHAMON, Pierre  
APPLICANT: KIENY, Marie-Paule  
APPLICANT: LAIHE, Richard  
APPLICANT: HAREUVENI, Mara  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P. O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22113-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,116  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/479,537  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1867 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1727  
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."  
OTHER INFORMATION:  
OTHER INFORMATION:  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."  
OTHER INFORMATION:  
OTHER INFORMATION:  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA or ACG; and Asn = AAT or AAC."  
OTHER INFORMATION:  
OTHER INFORMATION:  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."  
OTHER INFORMATION:  
OTHER INFORMATION:  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."  
OTHER INFORMATION:

Query Match	7.48%	Score 424.5	DB 4	Length 1867	
Best Local Similarity	23.78%	Pred. No. 4.5e-20			
Matches 255	Conservative 69	Mismatches 511	Indels 241	Gaps 53	
QY	75	TSLTVNKETTVETKEKTTTNKOTSTIDGKEKTTSAKETOSIEKTSAKD-LAPTSKVLAKPT	133		
Db	16	TVLTV-----VTGSGHASSTPGGCKETSATQGRSSVSPSEKKNAAVSTSSLSHS	65		
QY	134	P-KAETTVKGP--ALTPKEP-----TPPTTKEPASTPTKEPPTTJKSAP	176		
Db	66	PGSGSSSTTGOQDYLAPATEPAGSAAIWGQDVTSVPATRALGSLTPRAHDVT---SAP	122		
QY	177	TTKEPEAPPTT-----KSAPTTKEP-----APTTTKEPAPTTKEPAPT	217		
Db	123	--DNKRAPGSTAPXAHGVTISAPDXRXPGSTAPXAHGVTISAPDXKRPXGSTAPXAHGVTIS	180		
QY	218	TKE--PAPPTT-----KSAPTTKEPAPTTPK---KRAPTTTKEPAPTTPKEPPTT	264		
Db	181	APDXRXPXGSTAPXAHGVTISAPDXRXPSTAPXAHGVTISAPDXKRPXGSTAPXAHGVTIS	240		
QY	265	PKE--PAPTTKEP-----APTTKEPAPTTPK---KRAPTTTKEPAPTTPKEPAPT	311		
Db	241	APDXRXPXGSTAPXAHGVTISAPDXRXPSTAPXAHGVTISAPDXKRPXGSTAPXAHGVTIS	300		
QY	312	TKEPSPTTTPKEPAPT--TKSAPTTTKEPAPT-----TKSAPTTTPKPSPTTTK-----E	360		
Db	301	APDXRP--XPGSTAPXAHGVTISAPDXKRPXGSTAPXAHGVTISAPDXRXPXGSTAPXAHGVT	359		
QY	361	PAPTTTPKEPAPTTPK---KRAPTTTKEPAPTTPK---EPAPTTTTPKPAATTAKEPAPT	412		
Db	360	SAPDXRXPXGSTAPXAHGVTISAPDXKRPXGSTAPXAHGVTISAPDXKRPXGSTAPXAHGVT	419		
QY	413	TPKETAP---TPPKLTPPTPEKLAPTTTKEPAPTTPPELAPTTPEEPPTTTEEPAPT	468		
Db	420	SAPDXRXPXGSTAPXAHGVTIS---APDXKRPXGSTAPXAHGVTISAPDXRXP---PST	471		
QY	469	TPKA---AANTTKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPPKGT--APTTLKEPAP	523		
Db	472	APXAHGVTISAPDXRXPXGSTAPXAHGVTISAPDXRP--XPSTAPXAHGVTISAPDXKRPXGS	530		
QY	524	TPP-----KKPAPKELAPTTTKEPSTTSODKRAP--TPPKGAPTTTPKEPAPTTP	571		
Db	531	TPXAHGVTISAPDXKRPXGSTAP--XAHGVTISAPDXKRPXGSTAPXAHGVTIS---APDXR	585		
QY	572	KEPAPTTPKG---TAPTTLKEPAPTTP-----KKPAPKELAPTTTKEPSTSD	617		
Db	586	PXPGSTAPXAHGVTISAPDXRXPSTAPXAHGVTISAPDXKRPXGSTAP--XAHGVTISAPDX	644		
QY	618	KRAP--TPPK---ETAPTTPKEPAPTTP-----KKRAPTTPEPPTTSEVSTP	661		
Db	645	RXPXGSTAPXAHGVTISAPDXRXPXGSTAPXAHGVTISAPDXRXPXGS--TAPXAHGVTISAP	702		
QY	662	TTTTKEPPT-----IHSPSESTPELSAPPTKALENSKEEGCVTTTPPA-----	707		
Db	703	DXRXPXGSTAPXAHGVTISAPDXRXPSTAPXAHGVTISAPDXKRPXGSTAPXAHGVTISAP	762		
QY	708	--TKPEWTTTAKDK---TERDLRTTTPETTTAAAPKMTKETATTTTEKTTESKITAATTOVTS	763		
Db	763	DXRXPXGSTAPXAHGVTISAPDXRXPST--APXAHGVTISAPDXRXPXGSTAPXAHGVTIS	820		
QY	764	TTTTODTTPKITTTLTTLTLARVYTTTKTITTEIIMKPEPELAKKPDRAITNSKATTPKQ	823		
Db	821	APDXRXPXGP-----STAPXAHGVT-----SAPDXRXPSTAPXAHGVTISAPDXKRPAP-	868		
QY	824	KRPKAKKKTSTSKKKTTPRVKRPKTTPPTPKKMTSTPELNPSTSLAEMLQTTTPRNOT	883		
Db	869	--GSTAPXAHGVTISAPDXRP--XPGSTAPXAHGVTISAP--PDXRPX-----PGST	911		
QY	884	--PNSKILVEVNPSESDAGAGEPTPHMLLRPHVEMPEVTPMDYLLPRVPOGIIINPMLSD	942		
Db	912	APXAHGVTISAPDXRXPXGSTAPXAH-----GVTISAPDXKRPXGSTA---PXAHG	957		
QY	943	ETWICNGKVDGLTTL--LRNGTLVAPRGHYFWMLLSPFSPSPAPRITTEVWGIQPSID	997		

Db 958 VTSAPDXRXPXGSTAPXAHGVTSAPD-----PDXRXPXGSTAPXAHGVTSAPD 1003

RESULT 13

US-08-479-537A-2

Sequence 2, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMON, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LARHE, Richard

APPLICANT: HAREVENT, Maita

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Rodin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1899

OTHER INFORMATION: /note="The amino acids spanning

OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat

OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such

OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note="Amino acid 134 is XI - Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note="Amino acid 144 is Y - Xaa

OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note="Amino acid 147 is X2 - Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note="Amino acids 1 to 21 are a

OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 7.4%; Score 424.5; DB 2; Length 2035;

Best Local Similarity 23.7%; Pred No. Se-20;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

Db 75 TSLVNETTETVETKETTNNKQSTDGKETTSAKETOSIEKTSAKD-LAPTSVLAKPT 133

Db 16 TVLTV-----VTGSGHASTPGEKERSATORSVPSTKNAVSMTSVLSHS 65

Db 134 P-KAETTKGP--ALTTRKP-----TPTRKBPASTTRKEPTTTIKSAP 176

Db 66 PGGSGSTTGGODVTLAPATEPASGSAATWGODVTSVPTRALGSTTPPAHDVT---SAP 122

Db 177 TPKEPAPPTT-----KSAPTTKPEP-----APTTKKEPAPPTTPKEPAPT 217

Db 123 ---DNKPAEGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 180

Db 218 TKE--PAPTT-----KSAPTTKPEPAPPTPK---KPAPTTKPEPAPPTPKPTT 264

Db 181 APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 240

Db 265 PKE--PAPTTKPEP-----APTTRKPEPAPTPK---KPAPTTKPEPAPPTPKPTT 311

Db 241 APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 300

Db 312 TKEPPTPKPEPAPPT---TKSAPTTKPEPAPT-----TKSAPTTKPEPSTTTK---E 360

Db 301 APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 359

Db 361 PAPTTKPEPAPPTPK---KPAPTTKPEPAPPTPK---EPAPTTTKKPAPTTAPKEPAPT 412

Db 360 SAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 419

Db 413 TPKEPAP---TPTRKLLPTTPKEKLAPPTPKPAPPTPEELAPPTPEEPPTTPEEPAPT 468

Db 420 SAPDXRXPXGSTAPXAHGVTS---APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 471

Db 469 TPKA---AAPTTRKPEPAPPTPKPEPAPPTPKPEPAPPTPKPTTAPTTKBPAP 523

Db 472 APXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS---APDXR 585

Db 524 TTP-----KKPAKELAPTTTKEPTSTTSKRAP---TPTRKGTAPPTPKPEPAPT 571

Db 531 TAPXAHGVTSAPDXRXPXGSTAP---XAHGVTSAPDXRXPXGSTAPXAHGVTS---APDXR 585

Db 572 KEPAPTPKG---TAPTTTLEKAPPTP-----KKPAPLELAPTTTGGTSTTSD 617

Db 586 PXPSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDX 644

Db 618 KPAP--TTPK---ETAPTTKPEPAPPTP-----KKPAPTPETPTTSEVSTP 661

Db 645 RXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAP 702

Db 662 TTTEPTT-----HKSPDESTPELSAETPKALENSKEEGVPTTTKPA-- 707

Db 703 DXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAP 762

```

OY 708 -TRPEMTTAKD---TTERDLRTTPEYTTAAPKMTKETATTEKTESKITATTQVTS 763
DB 763 DAPRXPSTAPXAHGVTAPDXRXPXGSGT--APXAHGVTAPDXRXPXGSGTAPXAHGVT 820
OY 764 TTTQDTTPPKITTLKTTTLPVTTTKTITTEIMNKKEETAKRKDRATNSKATTPRQ 823
DB 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSGTAPXAHGVTAPDXRXPX 868
OY 824 KPTKAKKPTSTKPKTMRVRKPKTTPPKMTSTMPLENTSRIAEAMLQTTTRPNOT 883
DB 869 -GSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTSA-PDXRXP-----PGST 911
OY 884 -PNSKLVENPXSSEDAGAEGETPHMLRPHVFMPEVTPDMQTLPRVNOGIIINMLSD 942
DB 912 APXAHGVTAPDXRXPXGSGTAPXAH-----GVTAPDXRXPXGSGTAPXAHGVTAPDXRXP 957
OY 943 ETNICKGKPYDGLTT-LRNGTLVAFRGHYFWMLSPPSPSPARLITEVWGIPSPID 997
DB 958 VTSAPDXRXPXGSGTAPXAHGVTSA-----PDXRXPXGSGTAPXAHGVTAPDXRXP 1003

```

```

RESULT 14
US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMBER, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREJVENT, Maira
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TITL OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
City: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Testin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid

```

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repea
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1.21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match 7.4%; Score 424.5; DB 4; Length 2035;
Best Local Similarity 23.7%; Pred. No. 5e-20; Indels 241; Gaps 53;
Matches 255; Conservative 69; Mismatches 511;

OY 75 TSLTVNKETTVERKETTNTKSTGKERTTSARETOSIEKTSKD-LAPTSKYLAKPT 133
DB 16 TVLTIV-----VTGSGHASSSTPGGKEKETSATGROSSVPSSTENAVSMSSVLSHS 65
OY 134 P-KAETTRKGP--ALTTPKEP-----TPTPKKEPASTTPKEPTTITKSNP 176
DB 66 PGSGSSTGQGVTLAPATEPAGSAAATGODVTSVPTRPALGSTTPPAHDVT--SAP 122
OY 177 TPKEPAPPTT-----KSAPTPKEP-----APTTPKEPAPPTTPKEPAPT 217
DB 123 --DNKRAPGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVT 180
OY 218 TRE--PAPPTT-----KSAPTPKEPAPPTPK-----KPAPTPKEPAPPTTPKEPAPT 264
DB 181 APDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVT 240
OY 265 PKE--PAPTPKEP-----APTTPKEPAPPTAK-----KPAPTPKEPAPPTTPKEPAPT 311
DB 241 APDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVT 300
OY 312 TKESPPTTPKEPAPT--TKSAPTTTKEPAPT-----TKSAPTTTPKESPTTTK-----E 360
DB 301 APDXRXP--XPGSTAPXAHGVTAPDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVT 359
OY 361 PAPTPKEPAPPTPK-----KPAPTPKEPAPPTPK-----EPAPTTTKKRAPAPKAPAPPT 412
DB 360 SAPDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVT 419
OY 413 TPKEPTAP-----TPPKLITPTTPKEKLAFTTPKEKAPPTTPELATTTPEEPPTTPKEPAPT 468
DB 420 SAPDXRXPXGSGTAPXAHGVT-----APDXRXPXGSGTAPXAHGVTAPDXRXPXGSGTAPXAHGVT 471
OY 469 TPKA-----AAPNTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPTTPKEPAPT 523

```

```

Db 472 APXAHGVTSAADKRXPGSTAPXAHGVTSAADKRXPGSTAPXAHGVTSAADKRXPGS 530
QY 524 TTP-----KKPAKELAPTTTKEPTSTSDKAP--TTPKGAAPTTKEPAATTP 571
Db 531 TAPXAHGVTSAADKRXPGSTAP--XAHGVTSAADKRXPGSTAPXAHGVTSA--APDXR 585
QY 572 KEPAATTPKPG--TAPTTKEPAATTP-----KKPAKELAPTTTKEPTSTSD 617
Db 586 RXPGSTAPXAHGVTSAADKRXPGSTAPXAHGVTSAADKRXPGSTAP--XAHGVTSAADK 644
QY 618 KPAP--TTPK-----ETAPTTKEPAATTP-----KKPAATTPETTPPTTSEVSTP 661
Db 645 RXPGSTAPXAHGVTSAADKRXPGSTAPXAHGVTSAADKRXPGS--TAPXAHGVTSA 702
QY 662 TTPKEPT-----IHKSPDESTPELSAEPPTKALENSKEGCVTTTKTPAA----- 707
Db 703 DXAPXGSTAPXAHGVTSAADKRXPGSTAPXAHGVTSAADKRXPGSTAPXAHGVTSA 762
QY 708 -TTPKMTTAKDK--TTERDLRTPEPTTAAPKMTKETATTEKTESKITATTQVTS 763
Db 763 DXAPXGSTAPXAHGVTSAADKRXPGST--APXAHGVTSAADKRXPGSTAPXAHGVT 820
QY 764 TTPQDTPPEKITLKTTLAPKVTTTTKTTTTEIMNKPEETAKPKDRATNSKATTPK 823
Db 821 APXKRXPG--STAPXAHGVT-----SAPDXRXPGSTAPXAHGVTSAADKRXPG 868
QY 824 KPAPKPKTSTTKPKTMRVRRKPTTPTRKWTSTMPLELNPSTSLAEMLQTTTPNOT 883
Db 869 -GSTAPXAHGVTSAADKRXPG--XPGSTAPXAHGVTSA--PDXRXPGST-----PGST 911
QY 884 -PASKLVEVNPSEDAAGEGETPHMLLRPHVFMPEVTPDMVLPRLVNOGIIINMLSD 942
Db 912 APXAHGVTSAADKRXPGSTAPXAHGVT-----GVTSAPXKRXPGSTAP--PXAHG 957
QY 943 ETNICKGRKPVDTLT--LRNGTLVAFRGHFMMLSPSPSPARITVEVGISPID 997
Db 958 VTSAPDXRXPGSTAPXAHGVTSA-----PDXRXPGSTAPXAHGVTSAAD 1003

RESULT 15
; Sequence 2, Application US/08276967
; Patent No. 5851817
; GENERAL INFORMATION:
; APPLICANT: Hardy, Daniel M.
; APPLICANT: Garbers, David L.
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,967
; FILING DATE: Submitted Herewith
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UTSD:418\KIT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679

```

```

; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-276-967-2

```

```

Query Match 7.4%; Score 419.5; DB 2; Length 2476;
Best Local Similarity 32.5%; Pred. No. 1,3e-19;
Matches 163; Conservative 43; Mismatches 177; Indels 119; Gaps 31;

```

```

QY 406 KEPAATTPKEPAATTPKKTTPTEPKLAPTTPEKPAATTPPEELAPTTPEEPPTPEEP 465
Db 313 PSETSVSTEKPAAPTE--KPLVSEIYITPEKVMIMKEPIVHT--EKPTVP--EKP 365
QY 466 APPTPKAAPNTPKEPA--PTTKREPAPTPKPEAPTTPEKETAATTPKGTAPTTLKEP 521
Db 366 TTPTEKSTVPT--KKPTVKEPTLPPE--GPTVPAE--RPTTPECGAVPPEKG--PTVTE- 418
QY 522 APPTPKPAKELAPTTTKEPTSTSDKPAATTPPGTAPT--TPKBPAPTTKREPAATTPK 580
Db 419 -----WPSHTEKSTVHTKELIPLTGKSTIPTKEPMVPTKRT-----TTP- 458
QY 581 GAAPTTLKEPAATTPKPKPAKELAPT--TKGPTSTT--SDKPAATTPKETAATTPKPEPA 636
Db 459 -TERTTIPAEKPTV--PIEKPMVPTERTTIPTEKTTIPTKPAVPTKELIVPT--EKPI 512
QY 637 PTPPKPAATTPETPPP-----TSEVSTPT-----TKKEPTTIKSPDESTPELSAEP 686
Db 513 VTEKPIVPTKELNITPEKLVLTERTTTPERTTIPTEKPIVPEKPSVPT--EKPIVPT 571
QY 687 PKALENSKEPGVPTTKTPAATKPEMTTAKDTERDLRTTPEPTTAAPKMTKETATTT 746
Db 572 -----EPTIIPTEKLVPT--ERTTTPKRTTTPITRTTPTI-----RTTTP 613
QY 747 EKTESKITATTTQVTSSTTODTTPFKITTLKTTTLABKVVTKKTTTTEIMNKPEETA 806
Db 614 ERTT-----TPTITTTTPTERTT--IPTKKT--VPEKTTIPT-----ERTI 652
QY 807 KPKDRATNSKATTPKPKOKPTAPKPKPTSKKPKTMPRVRKKTPTPKKMTSTMPLELNP 866
Db 653 AP-----TTPQP--SPTLVPTQPAVVMPEST-----SATVTPPTTTIASCP---PN 693
QY 867 SRIAEAMLQTTTRPNQTPNSKL 888
Db 694 AHFERCACPVSCQ--SPTPNCEL 714

```

```

Search completed: April 26, 2002, 16:17:20
Job time: 231 sec

```

